(12)

EUROPEAN PATENT APPLICATION

(43) Date of publication: 12.04.2000 Bulletin 2000/15

(21) Application number: 99119184.2

(22) Date of filing: 07.10.1999

(51) Int CI.7: **C12N 15/12**, C07K 14/75, C07K 14/78, C07K 14/51, C07K 14/47, C07K 14/495, C12P 21/02

(84) Designated Contracting States:

AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU

MC NL PT SE
Designated Extension States:
AL LT LV MK RO SI

(30) Priority: 09.10.1998 US 169768

(71) Applicant: United States Surgical Corporation Norwalk, Connecticut 06856 (US)

(72) Inventors:

Gruskin, Elliott A.
 Killingworth, CT 06419 (US)

 Buechter, Douglas D. Wallingford, CT 06492 (US)

 Zhang, Guanghui Guilford, CT 06473 (US)

 Connolly, Kevin Los Angeles, CA 90066 (US)

(74) Representative: HOFFMANN - EITLE Patent- und Rechtsanwälte Arabellastrasse 4 81925 München (DE)

(54) Extracellular matrix proteins with modified amino acid

(57) Incorporation of certain amino acid analogs into polypeptides produced by cells which do not ordinarily provide polypeptides containing such amino acid analogs is accomplished by subjecting the cells to growth media containing such amino acid analogs. The degree of incorporation can be regulated by adjusting the concentration of amino acid analogs in the media and/or by adjusting osmolality of the media. Such incorporation allows the chemical and physical characteristics of

polypeptides to be altered and studied. In addition, nucleic acid and corresponding proteins including a domain from a physiologically active peptide and a domain from an extracellular matrix protein which is capable of providing a self-aggregate are provided. Human extracellular matrix proteins capable of providing a self-aggregate collagen are provided which are produced by prokaryotic cells. Preferred codon usage is employed to produce extracellular matrix proteins in prokaryotics.

0 992 586

EEST AVAILABLE COPY

Description

1.1

5

20

25

30

45

50

BACKGROUND

1. Technical Field

[0001] Engineered polypeptides and chimeric polypeptides having incorporated amino acids which enhance or otherwise modify properties of such polypeptides.

2. Description of Related Art

[0002] Genetic engineering allows polypeptide production to be transferred from one organism to another. In doing so, a portion of the production apparatus indigenous to an original host is transplanted into a recipient. Frequently, the original host has evolved certain unique processing pathways in association with polypeptide production which are not contained in or transferred to the recipient. For example, it is well known that mammalian cells incorporate a complex set of post-translational enzyme systems which impart unique characteristics to protein products of the systems. When a gene encoding a protein normally produced by mammalian cells is transferred into a bacterial or yeast cell, the protein may not be subjected to such post translational modification and the protein may not function as originally intended. [0003] Normally, the process of polypeptide or protein synthesis in living cells involves transcription of DNA into RNA and translation of RNA into protein. Three forms of RNA are involved in protein synthesis: messenger RNA (mRNA) carries genetic information to ribosomes made of ribosomal RNA (rRNA) while transfer RNA (tRNA) links to free amino acids in the cell pool. Amino acid/tRNA complexes line up next to codons of mRNA, with actual recognition and binding being mediated by tRNA. Cells can contain up to twenty amino acids which are combined and incorporated in sequences of varying permutations into proteins. Each amino acid is distinguished from the other nineteen amino acids and charged to tRNA by enzymes known as aminoacyl-tRNA synthetases. As a general rule, amino acid/tRNA complexes are quite specific and normally only a molecule with an exact stereochemical configuration is acted upon by a particular aminoacyl-tRNA synthetase.

[0004] In many living cells some amino acids are taken up from the surrounding environment and some are synthesized within the cell from precursors, which in turn have been assimilated from outside the cell. In certain instances, a cell is auxotrophic, i.e., it requires a specific growth substance beyond the minimum required for normal metabolism and reproduction which it must obtain from the surrounding environment. Some auxotrophs depend upon the external environment to supply certain amino acids. This feature allows certain amino acid analogs to be incorporated into proteins produced by auxotrophs by taking advantage of relatively rare exceptions to the above rule regarding stere-ochemical specificity of aminoacyl-tRNA synthetases. For example, proline is such an exception, i.e., the amino acid activating enzymes responsible for the synthesis of prolyl-tRNA complex are not as specific as others. As a consequence certain proline analogs have been incorporated into bacterial, plant, and animal cell systems. See Tan et al., Proline Analogues Inhibit Human Skin Fibroblast Growth and Collagen Production in Culture, Journal of Investigative Dermatology, 80:261-267(1983).

[0005] A method of incorporating unnatural amino acids into proteins is described, e.g., in Noren et al., A General Method For Site-Specific Incorporation of Unnatural Amino Acids Into Proteins, Science, Vol. 244, pp. 182-188 (1989) wherein chemically acylated suppressor tRNA is used to insert an amino acid in response to a stop codon substituted for the codon encoding residue of interest. See also, Dougherty et al., Synthesis of a Genetically Engineered Repetitive Polypeptide Containing Periodic Selenomethionine Residues, Macromolecules, Vol. 26, No. 7, pp. 1779-1781 (1993), which describes subjecting an *E. coli* methionine auxotroph to selenomethionine containing medium and postulates on the basis of experimental data that selenomethionine may completely replace methionine in all proteins produced by the cell.

[0006] cis-Hydroxy-L-proline has been used to study its effects on collagen by incorporation into eukaryotic cells such as cultured normal skin fibroblasts (see Tan et al., supra) and tendon cells from chick embryos (see e.g., Uitto et al., Procollagen Polypeptides Containing cis-4-Hydroxy-L-proline are Overglycosylated and Secreted as Nonhelical Pro-γ-Chains, Archives of Biochemistry and Biophysics, 185:1:214-221(1978)). However, investigators found that trans-4-hydroxyproline would not link with proline specific tRNA of prokaryotic E. coli. See Papas et al., Analysis of the Amino Acid Binding to the Proline Transfer Ribonucleic Acid Synthetase of Escherichia coli, Journal of Biological Chemistry, 245:7:1588-1595(1970). Another unsuccessful attempt to incorporate trans-4-hydroxyproline into prokaryotes is described in Deming et al., In Vitro Incorporation of Proline Analogs into Artificial Proteins, Poly. Mater. Sci. Engin. Proceed., Vol. 71, p. 673-674 (1994). Deming et al. report surveying the potential for incorporation of certain proline analogs, i.e., L-azetidine-2-carboxylic acid, L-γ-thiaproline, 3,4-dehydroproline and L-trans-4-hydroxyproline into artificial proteins expressed in E. coli cells. Only L-azetidine-2-carboxylic acid, L-γ-thiaproline and 3,4 dehydroproline are reported as being incorporated into proteins in E. coli cells in vivo.

[0007] Extracellular matrix proteins ("EMPs") are found in spaces around or near cells of multicellular organisms and are typically fibrous proteins of two functional types: mainly structural, e.g., collagen and elastin, and mainly adhesive, e.g., fibronectin and laminin. Collagens are a family of fibrous proteins typically secreted by connective tissue cells. Twenty distinct collagen chains have been identified which assemble to form a total of about ten different collagen molecules. A general discussion of collagen is provided by Alberts, et al., The Cell, Garland Publishing, pp. 802-823 (1989), incorporated herein by reference. Other fibrous or filamentous proteins include Type I IF proteins, e.g., keratins; Type II IF proteins, e.g., vimentin, desmin and glial fibrillary acidic protein; Type III IF proteins, e.g., neurofilament proteins; and Type IV IF proteins, e.g., nuclear laminins.

5

10

15

20

- 25

30

35

50

[0008] Type I collagen is the most abundant form of the fibrillar, interstitial collagens and is the main component of the extracellular matrix. Collagen monomers consist of about 1000 amino acid residues in a repeating array of Gly-X-Y triplets. Approximately 35% of the X and Y positions are occupied by proline and *trans* 4-hydroxyproline. Collagen monomers associate into triple helices which consist of one $\alpha 2$ and two $\alpha 1$ chains. The triple helices associate into fibrils which are oriented into tight bundles. The bundles of collagen fibrils are further organized to form the scaffold for extracellular matrix.

[0009] In mammalian cells, post-translational modification of collagen contributes to its ultimate chemical and physical properties and includes proteolytic digestion of pro-regions, hydroxylation of lysine and proline, and glycosylation of hydroxylated lysine. The proteolytic digestion of collagen involves the cleavage of pro regions from the N and C termini. It is known that hydroxylation of proline is essential for the mechanical properties of collagen. Collagen with low levels of 4-hydroxyproline has poor mechanical properties, as highlighted by the sequelae associated with scurvy. 4-hydroxyproline adds stability to the triple helix through hydrogen bonding and through restricting rotation about C-N bonds in the polypeptide backbone. In the absence of a stable structure, naturally occurring cellular enzymes contribute to degrading the collagen polypeptide.

[0010] The structural attributes of Type I collagen along with its generally perceived biocompatability make it a desirable surgical implant material. Collagen is purified from bovine skin or tendon and used to fashion a variety of medical devices including hemostats, implantable gels, drug delivery vehicles and bone substitutes. However, when implanted into humans bovine collagen can cause acute and delayed immune responses.

[0011] As a consequence, researchers have attempted to produce human recombinant collagen with all of its structural attributes in commercial quantities through genetic engineering. Unfortunately, production of collagen by commercial mass producers of protein such as *E. coli* has not been successful. A major problem is the extensive post-translational modification of collagen by enzymes not present in *E. coli*. Failure of *E. coli* cells to provide proline hydroxylation of unhydroxylated collagen proline prevents manufacture of structurally sound collagen in commercial quantities

[0012] Another problem in attempting to use E. coli to produce human collagen is that E. coli prefer particular codons in the production of polypeptides. Although the genetic code is identical in both prokaryotic and eukaryotic organisms, the particular codon (of the several possible for most amino acids) that is most commonly utilized can vary widely between prokaryotes and eukaryotes. See, Wada, K.-N., Y. Wada, F. Ishibashi, T. Gojobori and T. Ikemura. Nucleic Acids Res. 20, Supplement: 2111-2118, 1992. Efficient expression of heterologous (e.g. mammalian) genes in prokaryotes such as E. coli can be adversely affected by the presence in the gene of codons infrequently used in E. coli and expression levels of the heterologous protein often rise when rare codons are replaced by more common ones. See, e.g., Williams, D.P., D. Regier, D. Akiyoshi, F. Genbauffe and J.R. Murphy, Nucleic Acids Res. 16: 10453-10467, 1988 and Höög, J.-O., H. v. Bahr-Lindström, H. Jörnvall and A. Holmgren, Gene, 43: 13-21, 1986. This phenomenon is thought to be related, at least in part, to the observation that a low frequency of occurrence of a particular codon correlates with a low cellular level of the transfer RNA for that codon. See, Ikemura, T.J. Mol. Biol. 158: 573-597, 1982 and Ikemura, T.J. Mol. Biol. 146: 1-21, 1981. Thus, the cellular tRNA level may limit the rate of translation of the codon and therefore influence the overall translation rate of the full-length protein. See, Ikemura, T.J. Mol. Biol. 146: 1-21, 1981; Bonekamp, F. and F.K. Jensen. Nucleic Acids Res. 16: 3013-3024, 1988; Misra, R. and P. Reeves, Eur. J. Biochem. 152: 151-155, 1985; and Post, L.E., G.D. Strycharz, M. Nomura, H. Lewis and P.P. Lewis. Proc. Natl. Acad. Sci. U.S.A. 76: 1697-1701, 1979. In support of this hypothesis is the observation that the genes for abundant E. coli proteins generally exhibit bias towards commonly used codons that represent highly abundant tRNAs. See, Ikemura, T.J. Mol. Biol. 146: 1-21, 1981; Bonekamp, F. and F.K. Jensen. Nucleic Acids Res. 16: 3013-3024, 1988; Misra, R. and P. Reeves, Eur. J. Biochem. 152: 151-155, 1985; and Post, L.E., G.D. Strycharz, M. Nomura, H. Lewis and P.P. Lewis. Proc. Natl. Acad. Sci. U.S.A. 76: 1697-1701, 1979. In addition to codon frequency, the codon context (i.e. the surrounding nucleotides) can also affect expression.

[0013] Although it would appear that substituting preferred codons for rare codons could be expected to increase expression of heterologous proteins in host organisms, such is not the case. Indeed, "it has not been possible to formulate general and unambiguous rules to predict whether the content of low-usage codons in a specific gene might adversely affect the efficiency of its expression in *E. coli.*" See page 524 of S.C. Makrides (1996), Strategies for Achieving High-Level Expression of Genes in *Escherichia coli.* Microbiological Reviews 60, 512-538. For example, in one

case, various gene fusions between yeast a factor and somatomedin C were made that differed only in coding sequence. In these experiments, no correlation was found between codon bias and expression levels in E. coli. Ernst, J.F. and Kawashima, E. (1988), J. Biotechnology, 7, 1-10. In another instance, it was shown that despite the higher frequency of optimal codons in a synthetic β -globin gene compared to the native sequence, no difference was found in the protein expression from these two constructs when they were placed behind the T7 promoter. Hernan et al. (1992), Biochemistry, 31, 8619-8628. Conversely, there are many examples of proteins with a relatively high percentage of rare codons that are well expressed in E. coli. A table listing some of these examples and a general discussion can be found in Makoff, A.J. et al. (1989), Nucleic Acids Research, 17, 10191-10202. In one case, introduction of non-optimal, rare arginine codons at the 3' end of a gene actually increased the yield of expressed protein. Gursky, Y.G. and Beabealashvilli, R.Sh. (1994), Gene 148, 15-21.

[0014] Failure to provide post-translational modifications such as hydroxylation of proline and the presence in human collagen of rare codons for *E. coli* may be contributing to the difficulties encountered in the expression of human collagen genes in *E. coli*.

SUMMARY

5

15

20

25

30

35

45

50

[0015] A method of incorporating an amino acid analog into a polypeptide produced by a cell is provided which includes providing a cell selected from the group consisting of prokaryotic cell and eukaryotic cell, providing growth media containing at least one amino acid analog selected from the group consisting of *trans*-4-hydroxyproline, 3-hydroxyproline, *cis*-4-fluoro-L-proline and combinations thereof and contacting the cell with the growth media wherein the at least one amino acid analog is assimilated into the cell and incorporated into at least one polypeptide.

[0016] Also provided is a method of substituting an amino acid analog of an amino acid in a polypeptide produced by a cell selected from the group consisting of prokaryotic cell and eukaryotic cell, which includes providing a cell selected from the group consisting of prokaryotic cell and eukaryotic cell, providing growth media containing at least one amino acid analog selected from the group consisting of trans-4-hydroxyproline, 3-hydroxyproline, cis-4-fluoro-L-proline and combinations thereof and contacting the cell with the growth media wherein the at least one amino acid analog is assimilated into the cell and incorporated as a substitution for at least one naturally occurring amino acid in at least one polypeptide.

[0017] A method of controlling the amount of an amino acid analog incorporated into a polypeptide is also provided which includes providing at least a first cell selected from the group consisting of prokaryotic cell and eukaryotic cell, providing a first growth media containing a first predetermined amount of at least one amino acid analog selected from the group consisting of *trans*-4-hydroxyproline, 3-hydroxyproline, *cis*-4-fluoro-L-proline and combinations thereof and contacting the first cell with the first growth media wherein a first amount of amino acid analog is assimilated into the first cell and incorporated into at least one polypeptide. At least a second cell selected from the group consisting of prokaryotic cell and eukaryotic cell, is also provided along with a second growth media containing a second predetermined amount of an amino acid analog selected from the group consisting of *trans*-4-hydroxyproline, 3-hydroxyproline, *cis*-4-fluoro-L-proline and combinations thereof and the at least second cell is contacted with the second growth media wherein a second amount of amino acid analog is assimilated into the second cell and incorporated into at least one polypeptide.

[0018] Also provided is a method of increasing stability of a recombinant polypeptide produced by a cell which includes providing a cell selected from the group consisting of prokaryotic cell and eukaryotic cell, and providing growth media containing an amino acid analog selected from the group consisting of trans-4-hydroxyproline, 3-hydroxyproline, cis-4-fluoro-L-proline and combinations thereof and contacting the cell with the growth media wherein the amino acid analog is assimilated into the cell and incorporated into a recombinant polypeptide, thereby stabilizing the polypeptide.

[0019] A method of increasing uptake of an amino acid analog into a cell and causing formation of an amino acid analog/tRNA complex is also provided which includes providing a cell selected from the group consisting of prokaryotic cell and eukaryotic cell, providing hypertonic growth media containing amino acid analog selected from the group consisting of *trans*-4-hydroxyproline, 3-hydroxyproline, *cis*-4-fluoro-L-proline and combinations thereof and contacting the cell with the hypertonic growth media wherein the amino acid analog is assimilated into the cell and incorporated into an amino acid analog/tRNA complex. In any of the other above methods, a hypertonic growth media can optionally be incorporated to increase uptake of an aminoacid analog into a cell.

[0020] A composition is provided which includes a cell selected from the group consisting of prokaryotic cell and eukaryotic cell, and hypertonic media including an amino acid analog selected from the group consisting of *trans*-4-hydroxyproline, *cis*-4-fluoro-L-proline and combinations thereof.

[0021] Also provided is a method of producing an Extracellular Matrix Protein (EMP) or a fragment thereof capable of providing a self-aggregate in a cell which does not ordinarily hydroxylate proline which includes providing a nucleic acid sequence encoding the EMP or fragment thereof which has been optimized for expression in the cell by substitution of codons preferred by the cell for naturally occurring codons not preferred by the cell, incorporating the nucleic acid

sequence into the cell, providing hypertonic growth media containing at least one amino acid selected from the group consisting of *trans*-4-hydroxyproline and 3-hydroxyproline, and contacting the cell with the growth media wherein the at least one amino acid is assimilated into the cell and incorporated into the EMP or fragment thereof.

[0022] Nucleic acid encoding a chimeric protein is provided which includes a domain from a physiologically active peptide and a domain from an extracellular matrix protein (EMP) which is capable of providing a self-aggregate. The nucleic acid may be inserted into a cloning vector which can then be incorporated into a cell.

[0023] Also provided is a chimeric protein including a domain from a physiologically active peptide and a domain from an extracellular matrix protein (EMP) which is capable of providing a self aggregate.

[0024] Also provided is human collagen produced by a prokaryotic cell, the human collagen being capable of providing a self aggregate.

[0025] Also provided is nucleic acid encoding a human Extracellular Matrix Protein (EMP) wherein the codon usage in the nucleic acid sequence reflects preferred codon usage in a prokaryotic cell.

BRIEF DESCRIPTION OF THE DRAWINGS

5

5.

15

20

40

45

50

[0026] Figure 1 is a plasmid map illustrating pMAL-c2.

[0027] Figure 2 is a graphical representation of the concentration of intracellular hydroxyproline based upon concentration of *trans*-4-hydroxyproline in growth culture over time.

[0028] Figure 2A is a graphical representation of the concentration of intracellular hydroxyproline as a function of sodium chloride concentration.

[0029] Figures 3A and 3B depict a DNA sequence encoding human Type 1 (α_1) collagen (SEQ. ID. NO. 1).

[0030] Figure 4 is a plasmid map illustrating pHuCol.

[0031] Figure 5 depicts a DNA sequence encoding a fragment of human Type 1 (α_1) collagen (SEQ. ID. NO.2.).

[0032] Figure 6 is a plasmid map illustrating pHuCol-FI.

²⁵ [0033] Figure 7 depicts a DNA sequence encoding a collagen-like peptide wherein the region coding for gene collagen-like peptide is underlined (SEQ. ID. NO. 3).

[0034] Figure 8 depicts an amino acid sequence of a collagen-like peptide (SEQ. ID. NO. 4).

[0035] Figure 9 is a plasmid map illustrating pCLP.

[0036] Figure 10 depicts a DNA sequence encoding mature bone morphogenic protein (SEQ. ID. NO. 5).

30 [0037] Figure 11 is a plasmid map illustrating pCBC.

[0038] Figure 12 is a graphical representation of the percent incorporation of proline and *trans*-4-hydroxyproline into maltose binding protein under various conditions.

[0039] Figure 13 depicts a collagen I (α1)/BMP-2B chimeric amino acid sequence (SEQ. ID. NO. 6).

[0040] Figure 14A-14C depicts a collagen I (α1)/BMP-2B chimeric nucleotide sequence (SEQ. ID. NO. 7).

35 [0041] Figure 15 depicts a collagen I (α1)/TGF-β₁amino acid sequence (SEQ. ID. NO. 8).

[0042] Figure 16A-16C depict a collagen I (α 1)/TGF- β_1 nucleotide sequence (SEQ. ID. NO. 9). Lower case lettering indicates non-coding sequence.

[0043] Figures 17A-17B depict a collagen I (α1)/decorin amino acid sequence (SEQ. ID. NO. 10).

[0044] Figure 18 depicts a collagen I (α1)/decorin peptide amino acid sequence (SEQ. ID. NO.11).

[0045] Figures 19A-19D depict a collagen I (α1)/decorin nucleotide sequence (SEQ. ID. NO. 12).

[0046] Figures 20A-20C depict a collagen/decorin peptide nucleotide sequence (SEQ. ID. NO. 13). Lower case lettering indicates non-coding sequence.

[0047] Figure 21 depicts a pMal cloning vector and polylinker cloning site.

[0048] Figure 22 depicts a polylinker cloning site contained in the pMal cloning vector of Fig. 21 (SEQ. ID. NO. 14).

[0049] Figure 23 depicts a pMal cloning vector containing a BMP/collagen nucleotide chimeric construct.

[0050] Figure 24 depicts a pMal cloning vector containing a TGF- β_1 /collagen nucleotide chimeric construct.

[0051] Figure 25 depicts a pMal cloning vector containing a decorin/collagen nucleotide chimeric construct.

[0052] Figure 26 depicts a pMal cloning vector containing a decorin peptide/collagen nucleotide chimeric construct.

[0053] Figure 27A-27E depicts a human collagen Type I (α_1) nucleotide sequence (SEQ. ID. NO. 15) and corresponding amino acid sequence (SEQ. ID. NO. 16).

[0054] Figure 28 is a schematic diagram of the construction of the human collagen gene from synthetic oligonucleotides.

[0055] Figure 29 is a schematic depiction of the amino acid sequence of chimeric proteins GST-CoIECoI (SEQ. ID. NO. 17) and GST-D4 (SEQ. ID. NO. 18).

⁵⁵ [0056] Figure 30 is a Table depicting occurrence of four proline and four glycine codons in the human Collagen Type I (α₁) gene with optimized codon usage (CoIECoI).

[0057] Figure 31 depicts a gel reflecting expression and dependence of expression of GST-D4 on hydroxyproline.

[0058] Figure 32 depicts a gel showing expression of GST-D4 in hypertonic media.

- [0059] Figure 33 is a graph showing circular dichroism spectra of native and denatured D4 in neutral phosphate buffer.
- [0060] Figure 34 depicts a gel representing digestion of D4 with bovine pepsin.

15

25

- [0061] Figure 35 depicts a gel representing expression of GST-H Col and GST-ColECol under specified conditions.
- [0062] Figure 36 depicts a gel representing expression of GST-CM4 in media with or without Nacl and either proline or hydroxyproline.
 - [0063] Figure 37 depicts a gel of six hour post induction samples of GST-CM4 expressed in *E. coli* with varying concentrations of NaCl.
 - [0064] Figure 38 depicts a gel of 4 hour post induction samples of GST-CM4 expressed in *E. coli* with constant amounts of hydroxyproline and varying amounts of proline.
- [0065] Figures 39A-39E depict the nucleotide (SEQ. ID. NO. 19) and amino acid (SEQ. ID. NO. 20) sequence of HuCol^{Ec}, the helical region of human Type I (α₁) collagen plus 17 amino terminal extra-helical amino acids and 26 carboxy terminal extra-helical amino acids with codon usage optimized for E. coli.
 - [0066] Figure 40 depicts sequence and restriction maps of synthetic oligos used to reconstruct the first 243 base pairs of the human Type I (α_1) collagen gene with optimized *E. coli* codon usage. The synthetic oligos are labelled N1-1 (SEQ. ID. NO. 21), N1-2 (SEQ. ID. NO. 22), N1-3 (SEQ. ID. NO. 23) and N1-4 (SEQ. ID. NO. 24).
 - **[0067]** Figure 41 depicts a plasmid map of pBSN1-1 containing a 114 base pair fragment of human collagen Type I (α_1) with optimized *E. coli* codon usage.
 - [0068] Figure 42 depicts the nucleotide (SEQ. ID. NO. 25) and amino acid (SEQ. ID. NO. 26) sequence of a fragment of human collagen Type I (α_1) gene with optimized *E. coli* codon usage encoded by plasmid pBSN1-1.
- [0069] Figure 43 depicts a plasmid map of pBSN1-2 containing a 243 base pair fragment of human collagen Type I (α_1) with optimized *E. coli* codon usage.
 - [0070] Figure 44 depicts the nucleotide (SEQ. ID. NO. 27) and amino acid (SEQ. ID. NO. 28) sequence of a fragment of human collagen Type I (α_1) gene with optimized *E. coli* codon usage encoded by plasmid pBSN1-2.
 - [0071] Figure 45 depicts a plasmid map of pHuCol^{Ec} containing human collagen Type I (α_1) with optimized *E. coli* codon usage.
 - [0072] Figure 46 depicts a plasmid map of pTrc N1-2 containing a 234 nucleotide human collagen Type I (α_1) fragment with optimized *E. coli* codon usage.
 - [0073] Figure 47 depicts a plasmid map of pN1-3 containing a 360 nucleotide human collagen Type I (α_1) fragment with optimized *E. coli* codon usage.
- 30 [0074] Figure 48 depicts a plasmid map of pD4 containing a 657 nucleotide human collagen Type I (α₁) 3' fragment with optimized *E. coli* codon usage.
 - **[0075]** Figures 49A-49E depict the nucleotide (SEQ. ID. NO. 29) and amino acid (SEQ. ID. NO. 30) sequence of a helical region of human Type I (α_2) collagen plus 11 amino terminal extra-helical amino acids and 12 carboxy terminal extrahelical amino acids.
- [0076] Figures 50A-50E depict the nucleotide (SEQ. ID. NO. 31) and amino acid (SEQ. ID. NO. 32) sequence of $HuCol(\alpha_2)^{Ec}$, the helical region of human Type I (α_2) collagen plus 11 amino terminal extra-helical amino acids and 12 carboxy terminal extra-helical amino acids with codon usage optimized for *E. coli*.
 - [0077] Figure 51 depicts sequence and restriction maps of synthetic oligos used to reconstruct the first 240 base pairs of human Type I (α_2) collagen gene with optimized *E. coli* codon usage. The synthetic oligos are labelled N1-1 (α_2) (SEQ. ID. NO. 33), N1-2 (α_2) (SEQ. ID. NO. 34), N1-3 (α_2) (SEQ. ID. NO. 35) and N1-4 (α_2) (SEQ. ID. NO. 36).
 - [0078] Figure 52 depicts a plasmid map of pBsN1-I (α_2) containing a 117 base pair fragment of human collagen Type I (α_2) with optimized *E. coli* codon usage.
 - [0079] Figure 53 depicts a plasmid map of pBSN1-2 (α_2) containing a 240 base pair fragment of human collagen Type I (α_2) with optimized *E. coli* codon usage.
- [0080] Figure 54 depicts the nucleotide (SEQ. ID. NO. 37) and amino acid (SEQ. ID. NO. 38) sequence of a fragment of human collagen Type I (α_2) gene with optimized *E. coli* usage encoded by plasmid pBSN1-2(α_2).
 - [0081] Figure 55 depicts a plasmid map of pHucol(α_2)^{Ec} containing the entire human collagen Type I (α_2) gene with optimized *E. coli* codon usage.
 - **[0082]** Figure 56 depicts a plasmid map of pN1-2 (α_2) containing a 240 base pair fragment of human collagen Type I(α_2) with optimized *E. coli* codon usage.
 - [0083] Figure 57 depicts a gel reflecting expression of GST and TGF-β1 under specified conditions.
 - [0084] Figure 58 depicts a gel reflecting expression of MBP, FN-BMP-2A, FN-TGF-β1 and FN under specified conditions.
 - [0085] Figure 59 depicts a gel showing expression of GST-Coll under specified conditions.
- 55 [0086] Figure 60 depicts a plasmid map of pGST-CM4 containing the gene for glutathione S- transferase fused to the gene for collagen mimetic 4.
 - [0087] Figure 61 depicts the nucleotide (SEQ. ID. NO. 39) and amino acid (SEQ. ID. NO. 40) sequence of collagen mimetic 4.

[0088] Figure 62A depicts a chromatogram of the elution of hydroxyproline containing collagen mimetic 4 from a Poros RP2 column. The arrow indicates the peak containing hydroxyproline containing collagen mimetic 4.

[0089] Figure 62B depicts a chromatogram of the elution of proline-containing collagen mimetic 4 from a Poros RP2 column. The arrow indicates the peak containing proline containing collagen mimetic 4.

[0090] Figure 63A depicts a chromatogram of a proline amino acid standard (250 pmol).

5

15

20

25

35

45

50

55

- [0091] Figure 63B depicts a chromatogram of a hydroxyproline amino acid standard (250 pmol).
- [0092] Figure 63C depicts an amino acid analysis chromatogram of the hydrolysis of proline containing collagen mimetic 4.
- [0093] Figure 63D depicts an amino acid analysis chromatogram of the hydrolysis of hydroxyproline containing collagen mimetic 4.
- [0094] Figure 64 is a graph of OD600 versus time for cultures of *E. coli* JM109 (F-) grown to plateau and then supplemented with various amino acids.
- [0095] Figure 65 depicts a plasmid map of pcEc- α 1 containing the gene for HuCol(α 1)^{Ec}.
- [0096] Figure 66 depicts a plasmid map of pcEc- α 2 containing the gene for HuCol(α 2)^{Ec}
- [0097] Figure 67 depicts a plasmid map of pD4-α1 containing the gene for a 219 amino acid C-terminal fragment of Type I (α1) human collagen with optimized *E. coli* codon usage fused to the gene for glutathione S-transferase.
 - [0098] Figure 68 depicts a plasmid map of pD4- α 2 containing the gene for a 207 amino acid C-terminal fragment of Type I (α 2) human collagen with optimized *E. coli* codon usage fused to the gene for glutathione S-transferase.
 - [0099] Figure 69 depicts the predicted amino acid sequence from the DNA sequence of the first 13 amino acid acids of protein D4-α1 (SEQ. ID. NO. 41) and the amino acid sequence as experimentally determined (SEQ. ID NO. 42).
 - [0100] Figure 70 depicts the mass spectrum of hydroxyproline containing D4- α 1.
 - [0101] Figure 71 depicts the nucleotide sequence of a 657 nucleotide human collagen Type I (α1)3' fragment with optimized *E. coli* codon usage designated D4 (SEQ. ID. NO. 43).
- [0102] Figure 72 depicts the amino acid sequence of a 219 amino acid C-terminal fragment of human collagen Type I (α1) designed D4 (SEQ, ID, NO, 44).
 - [0103] Figure 73 is a plasmid map illustrating pGEX-4T. 1 containing the gene for glutatione S-transferase.
 - [0104] Figure 74 is a plasmid map illustrating pTrc-TGF containing the gene for the mature human TGF-β1 polypeptide.
 - [0105] Figure 75 is a plasmid map illustrating pTrc-Fn containing the gene for a 70 kDa fragment of human fibronectin.
- [0106] Figure 76 is a plasmid map illustrating pTrc-Fn-TGF containing the gene for a fusion protein of a 70 kDA fragment of human fibronectin and the mature human TGF-β1 polypeptide.
 - [0107] Figure 77 is a plasmid map illustrating pTrc-Fn-BMP containing the gene for a fusion protein of a 70 kDa fragment of human fibronectin and human bone morphogenic protein 2A.
 - [0108] Figure 78 is a plasmid map illustrating pGEX-HuColl^{Ec} containing the gene for a fusion between glutathione S-transferase and Type I (α 1) human collagen with optimized *E. coli* codon usage.
 - [0109] Figure 79 depicts the nucleotide sequence of a 627 nucleotide human collagen Type I (α 2) 3' fragment with optimized *E. coli* codon usage (SEQ. ID. NO.45).
 - [0110] Figure 80 depicts the amino acid sequence of a 209 amino acid C-terminal fragment of human collagen Type I (α 2) (SEQ. ID. NO. 46).
- 40 [0111] Figure 81 depicts the sequence of synthetic oligos used to reconstruct the first 282 base pairs of the gene for the carboxy terminal 219 amino acids of human Type I (α1) collagen with optimized E. coli codon usage designated N4-1 (SEQ. ID. NO. 47), N4-2 (SEQ. ID. NO. 48), N4-3 (SEQ. ID. NO. 49) and N4-4 (SEQ. ID. NO. 50).

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

[0112] Prokaryotic cells and eukaryotic cells can unexpectedly be made to assimilate and incorporate *trans*-4-hydroxyproline into proteins contrary to both Papas et al. and Deming et al., supra. Such assimilation and incorporation is especially useful when the structure and function of a polypeptide depends on post translational hydroxylation of proline not provided by the native protein production system of a recombinant host. Thus, prokaryotic bacteria such as *E. coli* and eukaryotic cells such as *Saccharomyces cerevisiae*, *Saccharomyces carlsbergensis* and *Schizosaccharomyces pombe* that ordinarily do not hydroxylate proline and additional eukaryotes such as insect cells including lepidopteran cell lines including *Spodoptera fiugiperda*, *Trichoplasia ni*, *Heliothis virescens*, *Bombyx mori* infected with a baculovirus; CHO cells, COS cells and NIH 3T3 cells which fail to adequately produce certain polypeptides whose structure and function depend on such hydroxylation can be made to produce polypeptides having hydroxylated prolines. Incorporation includes adding *trans*-4-hydroxyproline to a polypeptide, for example, by first changing an amino acid to proline, creating a new proline position that can in turn be substituted with *trans*-4-hydroxyproline or substituting a naturally occurring proline in a polypeptide with *trans*-4-hydroxyproline as well.

[0113] The process of producing recombinant polypeptides in mass producing organisms is well known. Replicable

expression vectors such as plasmids, viruses, cosmids and artificial chromosomes are commonly used to transport genes encoding desired proteins from one host to another. It is contemplated that any known method of cloning a gene, ligating the gene into an expression vector and transforming a host cell with such expression vector can be used in furtherance of the present disclosure.

[0114] Not only is incorporation of *trans*-4-hydroxyproline into polypeptides which depend upon *trans*-4-hydroxyproline for chemical and physical properties useful in production systems which do not have the appropriate systems for converting proline to *trans*-4-hydroxyproline, but useful as well in studying the structure and function of polypeptides which do not normally contain *trans*-4-hydroxyproline. It is contemplated that the following amino acid analogs may also be incorporated in accordance with the present disclosure: *trans*-4 hydroxyproline, 3-hydroxyproline, *cis*-4-fluoro-L-proline and combinations thereof (hereinafter referred to as the "amino acid analogs"). Use of prokaryotes and eukaryotes is desirable since they allow relatively inexpensive mass production of such polypeptides. It is contemplated that the amino acid analogs can be incorporated into any desired polypeptide. In a preferred embodiment the prokaryotic cells and eukaryotic cells are starved for proline by decreasing or eliminating the amount of proline in growth media prior to addition of an amino acid analog herein.

[0115] Expression vectors containing the gene for maltose binding protein (MBP), e.g., see Figure 1 illustrating plasmid pMAL-c2, commercially available from New England Bio-Labs, are transformed into prokaryotes such as *E. coli* proline auxotrophs or eukaryotes such as *S. cerevisiae* auxotrophs which depend upon externally supplied proline for protein synthesis and anabolism. Other preferred expression vectors for use in prokaryotes are commercially available plasmids which include pKK-223 (Pharmacia), pTRC (Invitrogen), pGEX (Pharmacia), pET (Novagen) and pQE (Quiagen). It should be understood that any suitable expression vector may be utilized by those with skill in the art.

15

25

45

55

[0116] Substitution of the amino acid analogs for proline in protein synthesis occurs since prolyl tRNA synthetase is sufficiently promiscuous to allow misacylation of proline tRNA with any one of the amino acid analogs. A sufficient quantity, i.e., typically ranging from about .001M to about 1.0 M, but more preferably from about .005M to about 0.5M of the amino acid analog(s) is added to the growth medium for the transformed cells to compete with proline in cellular uptake. After sufficient time, generally from about 30 minutes to about 24 hours or more, the amino acid analog(s) is assimilated by the cell and incorporated into protein synthetic pathways. As can be seen from Figures 2 and 2A, intracellular concentration of trans-4-hydroxyproline increases by increasing the concentration of sodium chloride in the growth media. In a preferred embodiment the prokaryotic cells and/or eukaryotic cells are starved for proline by decreasing or eliminating the amount of proline in growth media prior to addition of an amino acid analog herein.

[0117] Expression vectors containing the gene for human Type I (α1) collagen (DNA sequence illustrated in Figures 3 and 3A; plasmid map illustrated in Figure 4) are transformed into prokaryotic or eukaryotic proline auxotrophs which depend upon externally supplied proline for protein synthesis and anabolism. As above, substitution of the amino acid analog(s) occurs since prolyl tRNA synthetase is sufficiently promiscuous to allow misacylation of proline tRNA with the amino acid analog(s). The quantity of amino acid analog(s) in media given above is again applicable.

[0118] Expression vectors containing DNA encoding fragments of human Type 1 (α1) collagen (e.g., DNA sequence illustrated in Figure 5 and plasmid map illustrated in Figure 6) are transformed into prokaryotic or eukaryotic auxotrophs as above. Likewise, expression vectors containing DNA encoding collagen-like polypeptide (e.g., DNA sequence illustrated in Figure 7, amino acid sequence illustration in Figure 8 and plasmid map illustrated in Figure 9) can be used to transform prokaryotic or eukaryotic auxotrophs as above. Collagen-like peptides are those which contain at least partial homology with collagen and exhibit similar chemical and physical characteristics to collagen. Thus, collagen-like peptides consist, e.g., of repeating arrays of Gly-X-Y triplets in which about 35% of the X and Y positions are occupied by proline and 4-hydroxyproline. Collagen-like peptides are interchangeably referred to herein as collagen-like proteins, collagen-like polypeptides, collagen mimetic polypeptides and collagen mimetic. Certain preferred collagen fragments and collagen-like peptides in accordance herewith are capable of assembling into an extracellular matrix. In both collagen fragments and collagen-like peptides as described above, substitution with amino acid analog(s) occurs since prolyl tRNA synthetase is sufficiently promiscuous to allow misacylation of proline tRNA with one or more of the amino acid analog(s). The quantity of amino acid analog(s) given above is again applicable.

[0119] It is contemplated that any polypeptide having an extracellular matrix protein domain such as a collagen, collagen fragment or collagen-like peptide domain can be made to incorporate amino acid analog(s) in accordance with the disclosure herein. Such polypeptides include collagen, a collagen fragment or collagen-like peptide domain and a domain having a region incorporating one or more physiologically active agents such as glycoproteins, proteins, peptides and proteoglycans. As used herein, physiologically active agents exert control over or modify existing physiologic functions in living things. Physiologically active agents include hormones, growth factors, enzymes, ligands and receptors. Many active domains of physiologically active agents have been defined and isolated. It is contemplated that polypeptides having a collagen, collagen fragment or collagen-like peptide domain can also have a domain incorporating one or more physiologically active domains which are active fragments of such physiologically active agents. As used herein, physiologically active agent is meant to include entire peptides, polypeptides, proteins, glycoproteins, proteoglycans and active fragments of any of them. Thus, chimeric proteins are made to incorporate amino acid analog

(s) by transforming a prokaryotic proline auxotroph or a eukaryotic proline auxotroph with an appropriate expression vector and contacting the transformed auxotroph with growth media containing at least one of the amino acid analogs. For example, a chimeric collagen/bone morphogenic protein (BMP) construct or various chimeric collagen/growth factor constructs are useful in accordance herein. Such growth factors are well-known and include insulin-like growth factor, transforming growth factor, platelet derived growth factor and the like. Figure 10 illustrates DNA of BMP which can be fused to the 3' terminus of DNA encoding collagen, DNA encoding a collagen fragment or DNA encoding a collagen-like peptide. Figure 11 illustrates a map of plasmid pCBC containing a collagen/BMP construct. In a preferred embodiment, proteins having a collagen, collagen fragment or collagen-like peptide domain assemble or aggregate to form an extracellular matrix which can be used as a surgical implant. The property of self-aggregation as used herein includes the ability to form an aggregate with the same or similar molecules or to form an aggregate with different molecules that share the property of aggregation to form, e.g., a double or triple helix. An example of such aggregation is the structure of assembled collagen matrices.

[0120] Indeed, chimeric polypeptides which may also be referred to herein as chimeric proteins provide an integrated combination of a therapeutically active domain from a physiologically active agent and one or more EMP moieties. The EMP domain provides an integral vehicle for delivery of the therapeutically active moiety to a target site. The two domains are linked covalently by one or more peptide bonds contained in a linker region. As used herein, integrated or integral means characteristics which result from the covalent association of one or more domains of the chimeric proteins. The therapeutically active moieties disclosed herein are typically made of amino acids linked to form peptides, polypeptides, proteins, glycoproteins or proteoglycans. As used herein, peptide encompasses polypeptides and proteins

15

20

25

30

35

45

[0121] The inherent characteristics of EMPs are ideal for use as a vehicle for the therapeutic moiety. One such characteristic is the ability of the EMPs to form the self-aggregate. Examples of suitable EMPs are collagen, elastin, fibronectin, fibrinogen and fibrin. Fibrillar collagens (Type I, II and III) assemble into ordered polymers and often aggregate into larger bundles. Type IV collagen assembles into sheetlike meshworks. Elastin molecules form filaments and sheets in which the elastin molecules are highly cross-linked to one another to provide good elasticity and high tensile strength. The cross-linked, random-coiled structure of the fiber network allows it to stretch and recoil like a rubber band. Fibronectin is a large fibril forming glycoprotein, which, in one of its forms, consists of highly insoluble fibrils cross-linked to each other by disulfide bonds. Fibrin is an insoluble protein formed from fibrinogen by the proteolytic activity of thrombin during the normal clotting of blood.

[0122] The molecular and macromolecular morphology of the above EMPs defines networks or matrices to provide substratum or scaffolding in integral covalent association with the therapeutically active moiety. The networks or matrices formed by the EMP domain provide an environment particularly well suited for ingrowth of autologous cells involved in growth, repair and replacement of existing tissue. The integral therapeutically active moieties covalently bound within the networks or matrices provide maximum exposure of the active agents to their targets to elicit a desired response.

[0123] Implants formed of or from the present chimeric proteins provide sustained release activity in or at a desired locus or target site. Since it is linked to an EMP domain, the therapeutically active domain of the present chimeric protein is not free to separately diffuse or otherwise be transported away from the vehicle which carries it, absent cleavage of peptide bonds. Consequently, chimeric proteins herein provide an effective anchor for therapeutic activity which allows the activity to be confined to a target location for a prolonged duration. Because the supply of therapeutically active agent does not have to be replenished as often when compared to non-sustained release dosage forms, smaller amounts of therapeutically active agent may be used over the course of therapy. Consequently, certain advantages provided by the present chimeric proteins are a decrease or elimination of local and systemic side effects, less potentiation or reduction in therapeutic activity with chronic use, and minimization of drug accumulation in body tissue with chronic dosing.

[0124] Use of recombinant technology allows manufacturing of non-immunogenic chimeric proteins. The DNA encoding both the therapeutically active moiety and the EMP moiety should preferably be derived from the same species as the patient being treated to avoid an immunogenic reaction. For example, if the patient is human, the therapeutically active moiety as well as the EMP moiety is preferably derived from human DNA.

[0125] Osteogenic/EMP chimeric proteins provide biodegradable and biocompatible agents for inducing bone formation at a desired site. As stated above, in one embodiment, a BMP moiety is covalently linked with an EMP to form chimeric protein. The BMP moiety induces osteogenesis and the extracellular matrix protein moiety provides an integral substratum or scaffolding for the BMP moiety and cells which are involved in reconstruction and growth. Compositions containing the BMP/EMP chimeric protein provide effective sustained release delivery of the BMP moiety to desired target sites. The method of manufacturing such an osteogenic agent is efficient because the need for extra time consuming steps as purifying EMP and then admixing it with the purified BMP are eliminated. An added advantage of the BMP/EMP chimeric protein results from the stability created by the covalent bond between BMP and the EMP, i.e., the BMP portion is not free to separately diffuse away from the EMP, thus providing a more stable therapeutic agent.

[0126] Bone morphogenic proteins are class identified as BMP-1 through BMP-9. A preferred osteogenic protein for use in human patients is human BMP-2B. A BMP-2B/collagen IA chimeric protein is illustrated in Fig. 13 (SEQ. ID. NO. 6). The protein sequence illustrated in Fig. 15 (SEQ. ID. NO. 8) includes a collagen helical domain depicted at amino acids 1-1057 and a mature form of BMP-2B at amino acids 1060-1169. The physical properties of the chimeric protein are dominated in part by the EMP component. In the case of a collagen moiety, a concentrated solution of chimeric protein will have a gelatinous consistency that allows easy handling by the medical practitioner. The EMP moiety acts as a sequestering agent to prevent rapid desorption of the BMP moiety from the desired site and to provide sustained release of BMP activity. As a result, the BMP moiety remains at the desired site and provides sustained release of BMP activity at the desired site for a period of time necessary to effectively induce bone formation. The EMP moiety also provides a matrix which allows a patient's autologous cells, e.g., chondrocytes and the like, which are normally involved in osteogenesis to collect therein and form an autologous network for new tissue growth. The gelatinous consistency of the chimeric protein also provides a useful and convenient therapeutic manner for immobilizing active BMP on a suitable vehicle or implant for delivering the BMP moiety to a site where bone growth is desired.

Ġ.

15

20

25

30

45

50

[0127] The BMP moiety and the EMP moiety are optionally linked together by linker sequences of amino acids. Examples of linker sequences used are illustrated within the sequence depicted in Figs. 14A-14C (SEQ. ID. NO. 7), 16A-16C (SEQ. ID. NO. 9), 19A-19C (SEQ. ID. NO. 12) and 20A-20C (SEQ. ID. NO. 13), and are described in more detail below. Linker sequences may be chosen based on particular properties which they impart to the chimeric protein. For example, amino acid sequences such as Ile-Glu-Gly-Arg and Leu-Val-Pro-Arg are cleaved by factor XA and thrombin enzymes, respectively. Incorporating sequences which are cleaved by proteolytic enzymes into chimeric proteins herein provides cleavage at the linker site upon exposure to the appropriate enzyme and separation of the two domains into separate entities. It is contemplated that numerous linker sequences can be incorporated into any of the chimeric proteins.

[0128] In another embodiment, a chimeric DNA construct includes a gene encoding an osteogenic protein or a fragment thereof linked to gene encoding an EMP or a fragment thereof. The gene sequence for various BMPs are known, see, e.g., U.S. Patent Nos. 4,294,753, 4,761,471, 5,106,748, 5,187,076, 5,141,905, 5,108,922, 5,116,738 and 5,168,050, each incorporated herein by reference. A BMP-2B gene for use herein is synthesized by ligating oligonucleotides encoding a BMP protein. The oligonucleotides encoding BMP-2B are synthesized using an automated DNA synthesizer (Beckmen Oligo-1000). In preferred embodiment, the nucleotide sequence encoding the BMP is maximized for expression in *E. coli*. This is accomplished by using *E. coli* utilization tables to translate the sequence of amino acids of the BMP into codons that are utilized most often by *E. coli*. Alternatively, native DNA encoding BMP isolated from mammals including humans may be purified and used.

[0129] The BMP gene and the DNA sequence encoding an extracellular matrix protein are cloned by standard genetic engineering methods as described in Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor 1989, hereby incorporated by reference.

[0130] The DNA sequence corresponding to the helical and telepeptide region of collagen $I(\alpha 1)$ is cloned from a human fibroblast cell line. Two sets of polymerase chain reactions are carried out using cDNA prepared by standard methods from AG02261A cells. The first pair of PCR primers include a 5' primer bearing an XmnI linker sequence and a 3' primer bearing the BsmI site at nucleotide number 1722. The resulting PCR product consists of sequence from position 1 to 1722. The second pair of primers includes the BsmI site at 1722 and a linker sequence at the 3' end bearing a Bg1II site. The resulting PCR product consists of sequence from position 1722 to 3196. The complete sequence is assembled by standard cloning techniques. The two PCR products are ligated together at the BsmI site, and the combined clone is inserted into any vector with XmnI-Bg1II sites such as pMAL-c2 vector.

[0131] To clone the BMP-2B gene, total cellular RNA is isolated from human osteosarcoma cells (U-20S) by the method described by Robert E. Farrel Jr. (Academic-Press, CA, 1993 pp. 68-69) (herein incorporated by reference). The integrity of the RNA is verified by spectrophotometric analysis and electrophoresis through agarose gels. Typical yields of total RNA are 50 µg from a 100mm confluent tissue culture dish. The RNA is used to generate cDNA by reverse transcription using the Superscript pre-amplification system by Gibco BRL. The cDNA is used as template for PCR amplification using upstream and downstream primers specific for BMP-2B (GenBank HUMBMP2B accession #M22490). The resulting PCR product consists of BMP-2B sequence from position 1289-1619. The PCR product is resolved by electrophoresis through agarose gels, purified with gene clean (BIO 101) and ligated into pMal-c2 vector (New England Biolabs). The domain of human collagen I(\alpha1) chain is cloned in a similar manner. However, the total cellular RNA is isolated from a human fibroblast cell line (AG02261A human skin fibroblasts).

[0132] A chimeric BMP/EMP DNA construct is obtained by ligating a synthetic BMP gene to a DNA sequence encoding an EMP such as collagen, fibrinogen, fibrin, fibronectin, elastin or laminin. However, chimeric polypeptides herein are not limited to these particular proteins. Figs. 14A-14C (SEQ. ID. NO. 7) illustrate a DNA construct which encodes a BMP-2B/collagen I(al) chimeric protein. The coding sequence for an EMP may be ligated upstream and/or downstream and in-frame with a coding sequence for the BMP. The DNA encoding an EMP may be a portion of the gene or an entire EMP gene. Furthermore, two different EMPs may be ligated upstream and downstream from the BMP.

[0133] The BMP-2B/collagen I(al) chimeric protein illustrated in Figs. 14A-14C includes an XmnI linker sequence at base pairs (bp) 1-19, a collagen domain (bp 20-3190), a BgIII/BamHI linker sequence (bp 3191-3196), a mature form of BMP2b (bp 3197-3529) and a HindIII linker sequence (bp 3530-3535).

[0134] Any combination of growth factor and matrix protein sequences are contemplated including repeating units, or multiple arrays of each segment in any order.

[0135] Incorporation of fragments of both matrix and growth factor proteins is also contemplated. For example, in the case of collagen, only the helical domain may be included. Other matrix proteins have defined domains, such as laminin, which has EGF-like domains. In these cases, specific functionalities can be chosen to achieve desired effects. Moreover, it may be useful to combine domains from disparate matrix proteins, such as the helical region of collagen and the cell attachment regions of fibronectin. In the case of growth factors, specific segments have been shown to be removed from the mature protein by post translational processing. Chimeric proteins can be designed to include only the mature biologically active region. For example, in the case of BMP-2B only the final 110 amino acids are found in the active protein.

10

15

20

30

40

or downstream of alternate moieties.

[0136] In another embodiment, a transforming growth factor (TGF) moiety is covalently linked with an EMP to form a chimeric protein. The TGF moiety increases efficacy of the body's normal soft tissue repair response and also induces osteogenesis. Consequently, TGF/EMP chimeric proteins may be used for either or both functions. One of the fundamental properties of the TGF- β s is their ability to turn on various activities that result in the synthesis of new connective tissue. See, Piez and Sporn eds., Transforming Growth Factor- β s Chemistry, Biology and Therapeutics, Annals of the New York Academy of Sciences, Vol. 593, (1990). TGF- β is known to exist in at least five different isoforms. The DNA sequence for Human TGF- β_1 is known and has been cloned. See Derynck et al., Human Transforming Growth Factor-Beta cDNA Sequence and Expression in Tumour Cell Lines, Nature, Vol. 316, pp. 701-705 (1985), herein incorporated by reference. TGF- β_2 has been isolated from bovine bone, human glioblastoma cells and porcine platelets. TGF- β_3 has also been cloned. See ten Dijke, et al., Identification of a New Member of the Transforming Growth Factor- β Gene Family, Proc. Natl. Acad. Sci. (USA), Vol. 85, pp. 4715-4719 (1988) herein incorporated by reference.

[0137] A TGF-β/EMP chimeric protein incorporates the known activities of TGF-βs and provides integral scaffolding or substratum of the EMP as described above to yield a composition which further provides sustained release focal delivery at target sites.

101381 The TGF-β moiety and the EMP moiety are optionally linked together by linker sequences of amino acids.

Linker sequences may be chosen based upon particular properties which they impart to the chimeric protein. For example, amino acid sequences such as Ile-Glu-Glyn-Arg and Leu-Val-Pro-Arg are cleaved by Factor XA and Thrombin enzymes, respectively. Incorporating sequences which are cleaved by proteolytic enzymes into the chimeric protein provides cleavage at the linker site upon exposure to the appropriate enzyme and separation of the domains into separate entities. Fig. 15 depicts an amino acid sequence for a TGF-β₁/collagen IA chimeric protein (SEQ. ID. NO. 8). The illustrated amino acid sequence includes the collagen domain (1-1057) and a mature form of TGF β₁ (1060-1171). [0139] A chimeric DNA construct includes a gene encoding TGF-β₁ or a fragment thereof, or a gene encoding TGF- β_2 or a fragment thereof, or a gene encoding TGF- β_3 or a fragment thereof, ligated to a DNA sequence encoding an EMP protein such as collagen (I-IV), fibrin, fibrinogen, fibronectin, elastin or laminin. A preferred chimeric DNA construct combines DNA encoding TGF-β₁, a DNA linker sequence, and DNA encoding collagen IA. A chimeric DNA construct containing TGF- β_1 gene and a collagen I(α 1) gene is shown in Figs. 16A-16C (SEQ. ID. NO. 9). The illustrated construct includes an XmnI linker sequence (bp 1-19), DNA encoding a collagen domain (bp 20-3190), a BgIII linker sequence (bp 3191-3196), DNA encoding a mature form of TGF-β₁ (3197-3535), and an Xbal linker sequence (bp 3536-3541). [0140] The coding sequence for EMP may be ligated upstream and/or downstream and in-frame with a coding sequence for the TGFβ. The DNA encoding the extracellular matrix protein may encode a portion of a fragment of the EMP or may encode the entire EMP. Likewise, the DNA encoding the TGF-β may be one or more fragments thereof

[0141] In yet another embodiment, a dermatan sulfate proteoglycan moiety, also known as decorin or proteoglycan II, is covalently linked with an EMP to form a chimeric protein. Decorin is known to bind to type I collagen and thus affect fibril formation, and to inhibit the cell attachment-promoting activity of collagen and fibrinogen by binding to such molecules near their cell binding sites. Chimeric proteins which contain a decorin moiety act to reduce scarring of healing tissue. The primary structure of the core protein of decorin has been deduced from cloned cDNA. See Krusius et al., Primary Structure of an Extracellular Matrix Proteoglycan Core Protein-Deduced from Cloned cDNA, Proc. Natl. Acad. Sci. (USA), Vol. 83, pp. 7683-7687 (1986) incorporated herein by reference.

or the entire gene. Furthermore, two or more different TGF-\(\beta\)s or two or more different EMPs may be ligated upstream

[0142] A decorin/EMP chimeric protein incorporates the known activities of decorin and provides integral scaffolding or substratum of the EMP as described above to yield a composition which allows sustained release focal delivery to target sites. Figs. 17A-17B illustrate a decorin/collagen IA chimeric protein (SEQ. ID. NO. 10) in which the collagen domain includes amino acids 1-1057 and the decorin mature protein incudes amino acids 1060-1388. Fig. 18 illustrates a decorin peptide/collagen IA chimeric protein (SEQ. ID. NO. 11) in which the collagen helical domain includes amino

acids 1-1057 and the decorin peptide fragment includes amino acids 1060-1107. The decorin peptide fragment is composed of P46 to G93 of the mature form of decorin.

[0143] Further provided is a chimeric DNA construct which includes a gene encoding decorin or one or more fragments thereof, optionally ligated via a DNA linker sequence to a DNA sequence encoding an EMP such as collagen (I-IV), fibrin, fibrinogen, fibronectin, elastin or laminin. A preferred chimeric DNA construct combines DNA encoding decorin, a DNA linker sequence, and DNA encoding collagen I(α1). A chimeric DNA construct containing a decorin gene and a collagen I(α1) gene is shown in Figs. 19A-19D (SEQ. ID. NO. 12). The illustrated construct includes an XmnI linker sequence (bp 1-19), DNA encoding a collagen domain (bp 20-3190), a Bg1II linker sequence (bp 3191-3196), DNA encoding a mature form of decorin (bp 3197-4186) and a PstI linker sequence. A chimeric DNA construct containing a decorin peptide gene and a collagen I(α1) gene is shown in Figs. 20A-20C (SEQ. ID. NO. 13). The illustrated construct includes an XmnI linker sequence (bp 1-19), DNA encoding a collagen domain (bp 20-3190), a BgIII linker sequence (bp 3191-3196), DNA encoding a peptide fragment of decorin (bp 3197-3343), and a PstI linker sequence (bp 3344-3349).

[0144] The coding sequence for an EMP may be ligated upstream and/or downstream and in-frame with a coding sequence for decorin. The DNA encoding the EMP may encode a portion or fragment of the EMP or may encode the entire EMP. Likewise, the DNA encoding decorin may be a fragment thereof or the entire gene. Furthermore, two or more different EMPs may be ligated upstream and/or downstream from the DNA encoding decorin moiety.

15

25

30

45

[0145] Any of the above described chimeric DNA constructs may be incorporated into a suitable cloning vector. Fig. 21 depicts a pMal cloning vector containing a polylinker cloning site. Examples of cloning vectors are the plasmids pMal-p2 and pMal-c2 (commercially available from New England Biolabs). The desired chimeric DNA construct is incorporated into a polylinker sequence of the plasmid which contains certain useful restriction endonuclease sites which are depicted in Fig. 22 (SEQ. ID. NO. 14). The pMal-p2 polylinker sequence has XmnI, EcoRI, BamHI, HindIII, XbaI, Sa1I and PstI restriction endonuclease sites which are depicted in Fig 22. The polylinker sequence is digested with an appropriate restriction endonuclease and the chimeric construct is incorporated into the cloning vector by ligating it to the DNA sequences of the plasmid. The chimeric DNA construct may be joined to the plasmid by digesting the ends of the DNA construct and the plasmid with the same restriction endonuclease to generate "sticky ends" having 5' phosphate and 3' hydroxyl groups which allow the DNA construct to anneal to the cloning vector. Gaps between the inserted DNA construct and the plasmid are then sealed with DNA ligase. Other techniques for incorporating the DNA construct into plasmid DNA include blunt end ligation, poly(dA.dT) tailing techniques, and the use of chemically synthesized linkers. An alternative method for introducing the chimeric DNA construct into a cloning vector is to incorporate the DNA encoding the extracellular matrix protein into a cloning vector already containing a gene encoding a therapeutically active moiety.

[0146] The cloning sites in the above-identified polylinker site allow the cDNA for the collagen I(α 1)/BMP-2B chimeric protein illustrated in Figs. 14A-14C (SEQ. ID. NO. 7) to be inserted between the XmnI and the HindIII sites. The cDNA encoding the collagen I(α 1)/TGF- β 1 protein illustrated in Figs. 16A-16C (SEQ. ID. NO. 9) is inserted between the XmnI and the XbaI sites. The cDNA encoding the collagen I(α 1)/decorin protein illustrated in Figs. 19A-19D (SEQ. ID. NO. 12) inserted between the XmnI and the PstI sites. The cDNA encoding the collagen I(α 1)/decorin peptide illustrated in Figs. 20A-20C (SEQ. ID. NO. 13) is inserted between the XmnI and PstI sites.

[0147] Plasmids containing the chimeric DNA construct are identified by standard techniques such as gel electrophoresis. Procedures and materials for preparation of recombinant vectors, transformation of host cells with the vectors,
and host cell expression of polypeptides are described in Sambrook et al., Molecular Cloning: A Laboratory Manual,
supra. Generally, prokaryotic or eukaryotic host cells may be transformed with the recombinant DNA plasmids. Transformed host cells may be located through phenotypic selection genes of the cloning vector which provide resistance
to a particular antibiotic when the host cells are grown in a culture medium containing that antibiotic.

[0148] Transformed host cells are isolated and cultured to promote expression of the chimeric protein. The chimeric protein may then be isolated from the culture medium and purified by various methods such as dialysis, density gradient centrifugation, liquid column chromatography, isoelectric precipitation, solvent fractionation, and electrophoresis. However, purification of the chimeric protein by affinity chromatography is preferred whereby the chimeric protein is purified by ligating it to a binding protein and contacting it with a ligand or substrate to which the binding protein has a specific affinity.

[0149] In order to obtain more effective expression of mammalian or human eukaryotic genes in bacteria (prokaryotes), the mammalian or human gene may be placed under the control of a bacterial promoter. A protein fusion and purification system is employed to obtain the chimeric protein. Preferably, any of the above-described chimeric DNA constructs is cloned into a pMal vector at a site in the vector's polylinker sequence. As a result, the chimeric DNA construct is operably fused with the malE gene of the pMal vector. The malE gene encodes maltose binding protein (MBP). Fig. 23 depicts a pMal cloning vector containing a BMP/collagen DNA construct. A spacer sequence coding for 10 asparagine residues is located between the malE sequence and the polylinker sequence. This spacer sequence insulates MBP from the protein of interest. Figs. 24, 25 and 26 depict pMal cloning vectors containing DNA encoding

collagen chimeras with TGF- β_1 , decorin and a decorin peptide, respectively. The pMal vector containing any of the chimeric DNA constructs fused to the malE gene is transformed into *E. coli*.

5

15

20

25

30

35

45

to the chimeric protein. This technique utilizes the P_{tac} promoter of the pMal vector. The MBP contains a 26 amino acid N-terminal signal sequence which directs the MBP-chimeric protein through the *E. coli* cytoplasmic membrane. The protein can then be purified from the periplasm. Alternatively, the pMal-c2 cloning vector can be used with this protein fusion and purification system. The pMal-c2 vector contains an exact deletion of the malE signal sequence which results in cytoplasmic expression of the fusion protein. A crude cell extract containing the fusion protein is prepared and poured over a column of amylose resin. Since MBP has an affinity for the amylose it binds to the resin. Alternatively, the column can include any substrate for which MBP has a specific affinity. Unwanted proteins present in the crude extract are washed through the column. The MBP fused to the chimeric protein is eluted from the column with a neutral buffer containing maltose or other dilute solution of a desorbing agent for displacing the hybrid polypeptide. The purified MBP-chimeric protein is cleaved with a protease such as factor Xa protease to cleave the MBP from the chimeric protein. The pMal-p2 plasmid has a sequence encoding the recognition site for protease factor Xa which cleaves after the amino acid sequence Isoleucine-Glutamic acid-Glycine-Arginine of the polylinker sequence.

[0151] The chimeric protein is then separated from the cleaved MBP by passing the mixture over an amylose column. An alternative method for separating the MBP from the chimeric protein is by ion exchange chromatography. This system yields up to 100mg of MBP-chimeric protein per liter of culture. See Riggs, P., in Ausebel, F.M., Kingston, R. E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K. (eds.) Current Protocols in Molecular Biology, Supplement 19 (16.6.1-16.6.10) (1990) Green Associates/Wiley Interscience, New York, New England Biolabs (cat # 800-65S 9pMALc2) pMal protein fusion and purification system hereby incorporated herein by reference. (See also European Patent No. 286 239 herein incorporated by reference which discloses a similar method for production and purification of a protein such as collagen.)

[0152] Other protein fusion and purification systems may be employed to produce chimeric proteins. Prokaryotes such as *E. coli* are the preferred host cells for expression of the chimeric protein. However, systems which utilize eukaryote host cell lines are also acceptable such as yeast, human, mouse, rat, hamster, monkey, amphibian, insect, algae, and plant cell lines. For example, HeLa (human epithelial), 3T3 (mouse fibroblast), CHO (Chinese hamster ovary), and SP 2 (mouse plasma cell) are acceptable cell lines. The particular host cells that are chosen should be compatible with the particular cloning vector that is chosen.

[0153] Another acceptable protein expression system is the Baculovirus Expression System manufactured by Invitrogen of San Diego, California. Baculoviruses form prominent crystal occlusions within the nuclei of cells they infect. Each crystal occlusion consists of numerous virus particles enveloped in a protein called polyhedrin. In the baculovirus expression system, the native gene encoding polyhedrin is substituted with a DNA construct encoding a protein or peptide having a desired activity. The virus then produces large amounts of protein encoded by the foreign DNA construct. The preferred cloning vector for use with this system is pBlueBac III (obtained from Invitrogen of San Diego, California). The baculovirus system utilizes the *Autograph californica* multiple nuclear polyhidrosis virus (ACMNPV) regulated polyhedrin promoter to drive expression of foreign genes. The chimeric gene, i.e., the DNA construct encoding the chimeric protein, is inserted into the pBlueBac III vector immediately downstream from the baculovirus polyhedrin promoter.

[0154] The pBlueBac III transfer vector contains a B-galactosidase reporter gene which allows for identification of recombinant virus. The B-galactosidase gene is driven by the baculovirus ETL promoter (P_{ETL}) which is positioned in opposite orientation to the polyhedrin promoter (P_{PH}) and the multiple cloning site of the vector. Therefore, recombinant virus coexpresses B-galactosidase and the chimeric gene.

[0155] Spodoptera frugiperda (Sf9) insect cells are then cotransfected with wild type viral DNA and the pBlueBac III vector containing the chimeric gene. Recombination sequences in the pBlueBac III vector direct the vector's integration into the genome of the wild type baculovirus. Homologous recombination occurs resulting in replacement of the native polyhedrin gene of the baculovirus with the DNA construct encoding the chimeric protein. Wild type baculovirus which do not contain foreign DNA express the polyhedrin protein in the nuclei of the infected insect cells. However, the recombinants do not produce polyhedrin protein and do not produce viral occlusions. Instead, the recombinants produce the chimeric protein.

[0156] Alternative insect host cells for use with this expression system are Sf21 cell line derived from Spodoptera frugiperda and High Five cell lines derived from Trichoplusia ni.

[0157] Other acceptable cloning vectors include phages, cosmids or artificial chromosomes. For example, bacteriophage lambda is a useful cloning vector. This phage can accept pieces of foreign DNA up to about 20,000 base pairs
in length. The lambda phage genome is a linear double stranded DNA molecule with single stranded complementary
(cohesive) ends which can hybridize with each other when inside an infected host cell. The lambda DNA is cut with a
restriction endonuclease and the foreign DNA, e.g. the DNA to be cloned, is ligated to the phage DNA fragments. The
resulting recombinant molecule is then packaged into infective phage particles. Host cells are infected with the phage

particles containing the recombinant DNA. The phage DNA replicates in the host cell to produce many copies of the desired DNA sequence.

[0158] Cosmids are hybrid plasmid/bacteriophage vectors which can be used to clone DNA fragments of about 40,000 base pairs. Cosmids are plasmids which have one or more DNA sequences called "cos" sites derived from bacteriophage lambda for packaging lambda DNA into infective phage particles. Two cosmids are ligated to the DNA to be cloned. The resulting molecule is packaged into infective lambda phage particles and transfected into bacteria host cells. When the cosmids are inside the host cell they behave like plasmids and multiply under the control of a plasmid origin of replication. The origin of replication is a sequence of DNA which allows a plasmid to multiply within a host cell.

[0159] Yeast artificial chromosome vectors are similar to plasmids but allow for the incorporation of much larger DNA sequences of about 400,000 base pairs. The yeast artificial chromosomes contain sequences for replication in yeast. The yeast artificial chromosome containing the DNA to be cloned is transformed into yeast cells where it replicates thereby producing many copies of the desired DNA sequence. Where phage, cosmids, or yeast artificial chromosomes are employed as cloning vectors, expression of the chimeric protein may be obtained by culturing host cells that have been transfected or transformed with the cloning vector in a suitable culture medium.

10

15

20

35

45

[0160] Chimeric proteins disclosed herein are intended for use in treating mammals or other animals. The therapeutically active moieties described above, e.g., osteogenic agents such as BMPs, TGFs, decorin, and/or fragments of each of them, are all to be considered as being or having been derived from physiologically active agents for purposes of this description. The chimeric proteins and DNA constructs which incorporate a domain derived from one or more cellular physiologically active agents can be used for in vivo therapeutic treatment, in vitro research or for diagnostic purposes in general.

[0161] When used in vivo, formulations containing the present chimeric proteins may be placed in direct contact with viable tissue, including bone, to induce or enhance growth, repair and/or replacement of such tissue. This may be accomplished by applying a chimeric protein directly to a target site during surgery. It is contemplated that minimally invasive techniques such as endoscopy are to be used to apply a chimeric protein to a desired location. Formulations containing the chimeric proteins disclosed herein may consist solely of one or more chimeric proteins or may also incorporate one or more pharmaceutically acceptable adjuvants.

[0162] In an alternate embodiment, any of the above-described chimeric proteins may be contacted with, adhered to, or otherwise incorporated into an implant such as a drug delivery device or a prosthetic device. Chimeric proteins may be microencapsulated or macroencapsulated by liposomes or other membrane forming materials such as alginic acid derivatives prior to implantation and then implanted in the form of a pouchlike implant. The chimeric protein may be microencapsulated in structures in the form of spheres, aggregates of core material embedded in a continuum of wall material or capillary designs. Microencapsulation techniques are well known in the art and are described in the Encyclopedia of Polymer Science and Engineering, Vol. 9, pp. 724 et seq. (1980) hereby incorporated herein by reference.

[0163] Chimeric proteins may also be coated on or incorporated into medically useful materials such as meshes, pads, felts, dressings or prosthetic devices such as rods, pins, bone plates, artificial joints, artificial limbs or bone augmentation implants. The implants may, in part, be made of biocompatible materials such as glass, metal, ceramic, calcium phosphate or calcium carbonate based materials. Implants having biocompatible biomaterials are well known in the art and are all suitable for use herein. Implant biomaterials derived from natural sources such as protein fibers, polysaccharides, and treated naturally derived tissues are described in the Encyclopedia of Polymer Science and Engineering, Vol. 2, pp. 267 et seq. (1989) hereby incorporated herein by reference. Synthetic biocompatible polymers are well known in the art and are also suitable implant materials. Examples of suitable synthetic polymers include urethanes, olefins, terephthalates, acrylates, polyesters and the like. Other acceptable implant materials are biodegradable hydrogels or aggregations of closely packed particles such as polymethylmethacrylate beads with a polymerized hydroxyethyl methacrylate coating. See the Encyclopedia of Polymer Science and Engineering, Vol. 2, pp. 267 et seq. (1989) hereby incorporated herein by reference.

[0164] The chimeric protein herein provides a useful way for immobilizing or coating a physiologically active agent on a pharmaceutically acceptable vehicle to deliver the physiologically active agent to desired sites in viable tissue. Suitable vehicles include those made of bioabsorbable polymers, biocompatible nonabsorbable polymers, lactoner putty and plaster of Paris. Examples of suitable bioabsorbable and biocompatible polymers include homopolymers, copolymers and blends of hydroxyacids such as lactide and glycolide, other absorbable polymers which may be used alone or in combination with hydroxyacids including dioxanones, carbonates such as trimethylene carbonate, lactones such as caprolactone, polyoxyalkylenes, and oxylates. See the Encyclopedia of Polymer Science and Engineering, Vol. 2, pp. 230 et seq. (1989) hereby incorporated herein by reference.

[0165] These vehicles may be in the form of beads, particles, putty, coatings or film vehicles. Diffusional systems in which a core of chimeric protein is surrounded by a porous membrane layer are other acceptable vehicles.

[0166] In another aspect, the amount of amino acid analog(s) transport into a target cell can be regulated by con-

trolling the tonicity of the growth media. A hypertonic growth media increases uptake of *trans*-4-hydroxyproline into *E. coli* as illustrated in Figure 2A. All known methods of increasing osmolality of growth media are appropriate for use herein including addition of salts such as sodium chloride, KCl, MgCl₂ and the like, and sugars such as sucrose, glucose, maltose, etc. and polymers such as polyethylene glycol (PEG), dextran, cellulose, etc. and amino acids such as glycine. Increasing the osmolality of growth media results in greater intracellular concentration of amino acid analog (s) and a higher degree of complexation of amino acid analog(s) to tRNA. As a consequence, proteins produced by the cell achieve a higher degree of incorporation of amino acid analogs. Figure 12 illustrates percentage of incorporation of proline and hydroxyproline into MBP under isotonic and hypertonic media conditions in comparison to proline in native MBP. Thus, manipulating osmolality, in addition to adjusting concentration of amino acid analog(s) in growth media allows a dual-faceted approach to regulating their uptake into prokaryotic cells and eukaryotic cells as described above and consequent incorporation into target polypeptides.

[0167] Any growth media can be used herein including commercially available growth media such as M9 minimal medium (available from Gibco Life Technologies, Inc.), LB medium, NZCYM medium, terrific broth, SOB medium and others that are well known in the art.

15

20

25

30

35

40

45

50

55

[0168] Collagen from different tissues can contain different amounts of trans-4-hydroxyproline. For example, tissues that require greater strength such as bone contain a higher number of trans-4-hydroxyproline residues than collagen in tissues requiring less strength, e.g., skin. The present system provides a method of adjusting the amount of trans-4-hydroxyproline in collagen, collagen fragments, collagen-like peptides, and chimeric peptides having a collagen domain, collagen fragment domain or collagen-like peptide domain fused to a physiologically active domain, since by increasing or decreasing the concentration of trans-4-hydroxyproline in growth media, the amount of trans-4-hydroxyproline incorporated into such polypeptides is increased or decreased accordingly. The collagen, collagen fragments, collagen-like peptides and above-chimeric peptides can be expressed with predetermined levels of trans-4-hydroxyproline. In this manner physical characteristics of an extracellular matrix can be adjusted based upon requirements of end use. Without wishing to be bound by any particular theory, it is believed that incorporation of trans-4-hydroxyproline into the EMP moieties herein provides a basis for self aggregation as described herein.

[0169] In another aspect, the combination of incorporation of *trans*-4-hydroxyproline into collagen and fragments thereof using hyperosmotic media and genes which have been altered such that codon usage more closely reflects that found in *E. coli*, but retaining the amino acid sequence found in native human collagen, surprisingly resulted in production by *E. coli* of human collagen and fragments thereof which were capable of self aggregation.

[0170] The human collagen Type I (α_1) gene sequence (Figure 27A-27E) (SEQ. ID. NO. 15) contains a large number of glycine and proline codons (347 glycine and 240 proline codons) arranged in a highly repetitive manner. Table I below is a codon frequency tabulation for the human Type I (α_1) collagen gene. Of particular note is that the GGA glycine codon occurs 64 times and the CCC codon for proline occurs 93 times. Both of these codons are considered to be rare codons in *E. coli*. See, Sharp, P.M. and W.-H. Li. Nucleic Acids Res. 14: 7737-7749, 1986. These, and similar considerations for other human collagen genes are shown herein to account for the difficulty in expressing human collagen genes in *E. coli*.

TABLE 1

					1712	,					
Codon	Count	%age	Codon	Count	%age	Codon	Count	%age	Codon	Count	%age
TTT-	1	0.09	TCT-	18	1.70	TAT-	2	0.18	TGT-	0	0.00
Phe			Ser		l	Tyr			Cys		
TTC-	14	1.32	TCC-	4	0.37	TAC-	2	0.18	TGC-	0	0.00
Phe			Ser			Tyr			Cys		
TTA-	0	0.00	TCA-	2	0.18	TAA-	0	0.00	TGA-***	0	0.00
Leu			Ser			***					
TTG-	3	0.28	TCG-	0	0.00	TAG-	0	0.00	TGG-	0	0.00
Leu			Ser		1	***			Trp		
CTT-	4	0.37	ССТ-	141	13.33	CAT-	0	0.00	CGT-	26	2.45
Leu			Pro			His			Arg		
CTC-	7	0.66	CCC-	93	8.79	CÁC-	3	0.28	CGC-	6	0.56
Leu			Pro			His			Arg		
CTA-	0	0.00	CCA-	6	.0.56	CAA-	13	1.22	CGA-	11	1.04
Leu		!	Pro			Gln	,		Arg		
CTG-	7	0.66	CCG-	0	0.00	CAG-	17	1.60	CGG-	1	0.09
Leu	Ì		Pro			Gln			Arg		

TABLE 1 (continued)

5

∂1 10

15

20

25

30

45

50

55

Codon	Count	%age									
ATT-	6	0.56	ACT-	11	1.04	AAT-	6	0.56	AGT-	4	0.37
lle			Thr			Asn			Ser		
ATC-	0	0.00	ACC-	4	0.37	AAC-	5	0.47	AGC-	11	1.04
lle			Thr			Asn			Ser		
ATA-	1	0.09	ACA-	2	0.18	AAA-	19	1.79	AGA-	9	0.85
lle			Thr			Lys			Arg		
ATG-	7	0.66	ACG-	0	0.00	AAG-	19	1.79	AGG-	0	0.00
Met			Thr			Lys			Arg		
GTT-	10	0.94	GCT-	93	8.79	GAT-	23	2.17	GGT-	174	16.46
Val			Ala		1	Asp			Gly		
GTC-	5	0.47	GCC-	24	2.27	GAC-	11	1.04	GGC-	97	9.17
Val			Ala			Asp			Gty		
GTA-	0	0.00	GCA-	6	0.56	GAA-	24	2.27	GGA-	64	6.05
Val			Ala			Glu			Gly		
GTG-	5	0.47	GCG-	. 0	0.00	GAG-	25	2.36	GGG-	11	1.04
Val			Ala			Glu			Gly		

[0171] In a first step, the sequence of the heterologous collagen gene is changed to reflect the codon bias in E. coli as given in codon usage tables (e.g. Ausubel et al., (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York, New York; Wada et al., 1992, supra). Rare E. coli codons (See, Sharp, P.M. and W.-H. Li. Nucleic Acids Res. 14: 7737-7749, 1986) are avoided. Second, unique restriction enzyme sites are chosen that are located approximately every 120-150 base pairs in the sequence. In certain cases this entails altering the nucleotide sequence but does not change the amino acid sequence. Third, oligos of approximately 80 nucleotides are synthesized such that when two such oligos are annealed together and extended with a DNA polymerase they reconstruct a approximately 120-150 base pair section of the gene (Figure 28). The section of the gene encoding the very amino terminal portion of the protein has an initiating methionine (ATG) codon at the 5' end and a unique restriction site followed by a stop (TAAT) signal at the 3' end. The remaining sections have unique restriction sites at the 5' end and unique restriction sites followed by a TAAT stop signal the 3' end. The gene is assembled by sequential addition of each section to the preceding 5' section. In this manner, each successively larger section can be independently constructed and expressed. Figure 28 is a schematic representation of the construction of the human collagen gene starting from synthetic oligos. [0172] A fragment of the human Type I all collagen chain fused to the C-terminus of glutathione S-transferase (GST-D4, Fig. 29) (SEQ. ID. NO. 18) was prepared and tested for expression in E. coli strain JM109 (F-) under conditions of hyperosmotic shock. The collagen fragment included the C-terminal 193 amino acids of the triple helical region and the 26 amino acid C-terminal telopeptide. Fig. 29 is a schematic of the amino acid sequence of the GST-CoIECoI (SEQ. ID. NO. 17) and GST-D4 (SEQ. ID. NO. 18) fusion proteins. ColECol comprises the 17 amino acid N-terminal telopeptide, 338 Gly-X-Y repeating tripeptides, and the 26 amino acid C-terminal telopeptide. There is a unique methionine at the junction of GST and D4, followed by 64 Gly-X-Y repeats, and the 26 amino acid telopeptide. The residue (Phel99) in the C-terminal telopeptide of D4 where pepsin cleaves is indicated. The gene was synthesized for the collagen fragment from synthetic oligonucleotides designed to reflect optimal E. coli usage. Fig. 30 is a table depicting occurrence of the four proline and four glycine codons in the human Type I α1 gene (HCoI) and the Type I α1 gene with optimized E. coli codon usage (ColECol). Usage of the remaining codons in ColECol was also optimized for E. coli expression according to Wada et al., supra. Protein GST-D4 was efficiently expressed in JM109 (F-) in minimal media lacking proline but supplemented with Hyp and Nacl (See Figs. 31 and 32). Expression was dependent on induction with isopropyl-1-thio-β-galactopyranoside (IPTG), trans-4-hydroxyproline and NaCl. At a fixed Nacl concentration of 500 mM, expression was minimal at trans-4-hydroxyproline concentrations below ~20 mM while the expression level plateaued at trans-4-hydroxyproline concentrations above 40 mM. See Fig. 31 which depicts a gel showing expression and dependence of expression of GST-D4 on hydroxyproline. The concentration of hydroxyproline is indicated above each lane. Osmolyte (NaCl) was added at 500 mM in each culture and each was induced with 1.5 mM IPTG. The arrow marks the position of GST-D4. Likewise, at a fixed trans-4-hydroxyproline concentration of 40 mM, NaCl concentrations below 300 mM resulted in little protein accumulation and expression decreased above 700-800 mM NaCl. See Fig. 32 which depicts a gel showing expression of GST-D4 in hyperosmotic media. Lanes 2 and 3 are uninduced and induced samples, respectively, each without added osmolyte. The identity and quantity of osmolyte is indicated above each of the other lanes. Trans-4-Hydroxyproline was added at 40mM in each culture and all cultures except that in lane 1 were induced with 1.5 mM IPTG. The arrow marks the position of GST-D4.

5

10

15

20

30

35

ā.

45

55

[0173] Either sucrose or KC1 can be substituted for NaC1 as the osmolyte (See Fig. 32). Thus, the osmotic shock-mediated intracellular accumulation of trans-4-hydroxyproline was a critical determinant of expression rather than the precise chemical identity of the osmolyte. Despite the large number of prolines (66) in GST-D4, its size (46 kDA), and non-optimal growth conditions, it was expressed at \sim 10% of the total cellular protein. Expressed proteins of less than full-length indicative of aborted transcription, translation, or mRNA instability were not detected.

[0174] The gene for protein D4 contains 52 proline codons. In the expression experiments reflected in Figs. 31 and 32, it was expected that trans-4-hydroxyproline would be inserted at each of these codons resulting in a protein where trans-4-hydroxyproline had been substituted for all prolines. To confirm this, GST-D4 was cleaved with BrCN in 0.1 N HC1 at methionines within GST and at the unique methionine at the N-terminal end of D4, and D4 purified by reverse phase HPLC. Crude GST-D4 was dissolved in 0.1 M HC1 in a round bottom flask with stirring. Following addition of a 2-10 fold molar excess of clear, crystalline BrCN, the flask was evacuated and filled with nitrogen. Cleavage was allowed to proceed for 24 hours, at which time the solvent was removed in vacuo. The residue was dissolved in 0.1% trifluoroacetic acid (TFA) and purified by reverse-phase HPLC using a Vydac C4 RP-HPLC column (10 x 250 mm, 5 μ, 300 Å) on a BioCad Sprint system (Perceptive Biosystems, Framingham, MA). D4 was eluted with a gradient of 15 to 40% acetonitrile/0.1% TFA over a 45 min. period. D4 eluted as a single peak at 26% acetonitrile/0.1% TFA. Standard BrCN cleavage conditions (70% formic acid) resulted in extensive formylation of D4, presumably at the hydroxyl groups of the trans-4-hydroxyproline residues. Formylation of BrCN/formic acid-cleaved proteins had been noted before (Beavis et al., Anal. Chem., 62, 1836 (1990)). Amino acid analysis was carried out on a Beckman ion exchange instrument with post-column derivatization. N-terminal sequencing was performed on an Applied Biosystems sequencer equipped with an on-line HLPC system. Electrospray mass spectra were obtained with a VG Biotech BIO-Q quadropole analyzer by M-Scan, Inc. (West Chester, PA). For CD thermal melts, the temperature was raised in 0.5°C increments from 4°C to 85°C with a four minute equilibration between steps. Data were recorded at 221.5 nm. The thermal transition was calculated using the program ThermoDyne (MORE). The electrospray mass spectroscopy of this protein gave a single molecular ion corresponding to a mass of 20,807 Da. This mass is within 0.05% of that expected for D4 if it contains 100% trans-4-hydroxyproline in lieu of proline. Proline was not detected in amino acid analysis of purified D4, again consistent with complete substitution of trans-4-hydroxyproline for proline. To confirm further that trans-4-hydroxyproline substitution had only occurred at proline codons, the N-terminal 13 amino acids of D4 was sequenced as above. The first 13 codons of D4 specify the protein sequence H₂N-Gly-Pro-Pro-Gly-Leu-Ala-Gly-Pro-Pro-Gly-Glu-Ser-Gly (SEQ. ID. NO. 41). The sequence found was H₂N-Gly-Hyp-Hyp-Gly-Leu-Ala-Gly-Hyp-Hyp-Gly-Glu-Ser-Gly (SEQ. ID. NO. 42), see Fig. 69. Taken together, these results indicate that trans-4-hydroxyproline (Hyp) was inserted only at proline codons and that the fidelity of the E. coli translational machinery was not otherwise altered by either the high intracellular concentration or trans-4-hydroxyproline or hyperosmotic culture conditions.

[0175] To determine whether D4, containing trans-4-hydroxyproline in both the X and Y positions, forms homotrimeric helices and to compare stability to native collagen, the following was noted: In neutral pH phosphate buffer, D4 exhibits a circular dichroism (CD) spectrum characteristic of a triple helix (See Fig. 33 and Bhatnagar et al., Circular Dichroism and the Conformational Analysis of Biomolecules, G.D. Fasman, Ed. Plenum Press, New York, (1996 p. 183). Fig. 33 illustrates circular dichroism spectra of native and heat-denatured D4 in neutral phosphate buffer. HPLC-purified D4 was dissolved in 0.1M sodium phosphate, pH 7.0, to a final concentration of 1 mg/mL (E²⁸⁰=3628 M⁻¹·cm⁻¹). The solution was incubated at 4°C for two days to allow triple helices to form prior to analysis. Spectra were obtained on an Aviv model 62DS spectropolarimeter (Yale University, Molecular Biophysics and Biochemistry Department). A 1 mm path length quartz suprasil fluorimeter cell was used. Following a 10 min. incubation period at 4°C, standard wavelength spectra were recorded from 260 to 190 nm using 10 sec acquisition times and 0.5 nm scan steps. This spectrum is characterized by a negative ellipticity at 198 nm and a positive ellipticity at 221 nm. The magnitudes of both of these absorbances was greater in neutral pH buffer compared to acidic conditions. Comparable dependence of stability on pH has been noted for collagen-like triple helices. See, e.g., Venugopal et al., Biochemistry, 33, 7948 (1994). Heating at 85°C for five minutes prior to obtaining the CD spectrum decreased the magnitude of the absorbance at 198 nm and abolished the absorbance at 221 nm (Fig. 33). This behavior is also typical of the triple helical structure of collagen. See, R.S. Bhatnagar et al., Circular Dichroism and the Conformational Analysis of Biomolecules G.D. Fasman, Ed., supra. A thermal melt profile of D4 conducted as above in phosphate buffer gave a melting temperature of about 29°C. A fragment of the C-terminal region of the bovine Type I α1 collagen chain comparable in length to D4 forms homotrimeric helices with a melting temperature of 26°C. (See, A. Rossi, et al., Biochemistry 35, 6048 (1996)). [0176] Resistance to pepsin digestion is a second commonly used indication of triple helical structure. At 4°C, the majority of D4 is digested rapidly by pepsin to a protein of slightly lower molecular weight. Fig. 34 is a gel illustrating the result of digestion of D4 with bovine pepsin. Purified D4 was dissolved in 0.1 M sodium phosphate, pH 7.0, to 1.6 μg/μl and incubated at 4°C for 7 days. Aliquots (10 μl) were placed into 1.5 ml centrifuge tubes and adjusted with water and 1 M acetic acid solutions to 25 µl final volume and 200 mM final acetic acid concentration. Each tube was then incubated for 20 min. at the indicated temperature and pepsin (0.5 µl of a 0.25 µg/µl solution) was added to each tube

and digestion allowed to proceed for 45 minutes. Following digestion, samples were quenched with loading buffer and analyzed by SDS-PAGE. However, the initial pepsin cleavage product is resistant to further digestion up to ~30°C. Amino terminal sequencing as above of the initial pepsin cleavage product showed that the N-terminus was identical to that of full-length D4. Mass spectral analysis as above of the digestion product gave a parent ion with a molecular weight consistent with cleavage in the C-terminal telopeptide on the N-terminal side of Phe119 (See Fig. 29) suggesting that this portion of the protein is either globular or of ill-defined structure and rapidly cleaved by pepsin while the triple helical region is resistant to digestion. Thus, despite global *trans*-4-hydroxyproline for proline substitution in both the X and Y positions, D4 formed triple helices of stability similar to comparably sized fragments of bovine collagen containing Hyp at the normal percentage and only in the Y position.

10

15

20

25

30

35

40

50

[0177] The full-length human Type I α1 collagen chain, although more than four times the size of D4, also expressed as a N-terminal fusion with GST (GST-ColECol, Fig. 29) in JM109(F-) in Hyp/NaCl media. Fig. 35 is a gel depicting expression of GST-HCol and GST-ColECol. *Trans*-4-hydroxyproline was added at 40 mM and NaCl at 500 mM. Expression was induced with 1.5 mM IPTG. The arrow marks the position of GST-ColECol. In the procedures resulting in the gels shown in Figs. 31, 32 and 35, five ml cultures of JM109 (F-) harboring the expression plasmid in LB media containing 100 μg/ml ampicillin were grown overnight. Cultures were centrifuged and the cell pellets washed twice with five ml of M9/Amp media (See, J. Sambrook, E.F. Fritsch, T. Maniatis, *Molecular Cloning: A Laboratory Manual.* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989)) supplemented with 0.5% glucose and 100 μg/ml of all amino acids except glycine and alanine which were at 200 μg/ml and containing no proline. The cells were finally resuspended in five ml of the above media. Following incubation at 37°C for 30 min., hydroxyproline, osmolyte, or IPTG were added as indicated. After four hours, aliquots of the cultures were analyzed by SDS-PAGE.

[0178] Like D4, the gene for protein ColECol was constructed from synthetic oligonucleotides designed to mimic codon usage in highly-expressed E. coli genes. In contrast to GST-ColECol, expression from a GST-human Type I α1 gene fusion (pHCoI) identical to GST-CoIECoI in coded amino acid sequence but containing the human codon distribution could not be detected in Coomassie blue-stained SDS-PAGE gels of total cell lysates of induced JM109 (F-)/ pHCol cultures (Fig. 35). The gene for the Type I α1 collagen polypeptide was cloned by polymerase chain reaction of the gene from mRNA isolated from human foreskin cells (HS27, ATCC 1634) with primers designed from the published gene sequence (GenBank Z74615). The 5' primer added a flanking EcoR I recognition site and the 3' primer a flanking Hind III recognition site. The gene was cloned into the EcoR I/Hind III site of plasmid pBSKS+ (Stratagene, La Jolla, CA), four mutations corrected using the ExSite mutagenesis kit (Stratagene, La Jolla, CA), the sequence confirmed by dideoxy sequencing, and finally the EcoR I/Xho I fragment subcloned into plasmid pGEX-4T.1 (Pharmacia, Piscataway, NJ). The GST-HCol gene is expression-competent because a protein of the same molecular weight as GST-ColECol is detected when immunoblots of total cell lysates are probed with an anti-Type I collagen antibody. Thus, sequence or structural differences between the genes for ColECol and HCol are critical determinants of expression efficiency in E. coli. This is likely due to the codon distribution in these genes and ultimately to differences in tRNA isoacceptor levels in E. coli compared to humans. GST-ColECol, GST-D4, and GST-HCol do not accumulate in hyperosmotic shock media when proline is substituted for hydroxyproline or in rich media. A possible explanation is that the trans-4-hydroxyproline-containing proteins may be resistant to degradation because they fold into a protease-resistant triple helix while the proline-containing proteins do not adopt this structure. The large number of codons non-optimal for E. coli found in the human gene and the instability of proline-containing collagen in E. coli may, in part, explain why expression of human collagen in E. coli has not been previously reported.

[0179] As discussed above, collagen mimetic polypeptides, i.e., engineered polypeptides having certain compositional and structural traits in common with collagen are also provided herein. Such collagen mimetic polypeptides may also be made to incorporate amino acid analogs as described above. GST-CM4 consists of glutathione S-transferase fused to 30 repeats of a Gly-X-Y sequence. The Gly-X-Y repeating section mimics the Gly-X-Y repeating unit of human collagen and is referred to as collagen mimetic 4 or CM4 herein. Thus, the hydroxyproline-incorporating technology was also demonstrated to work with a protein and DNA sequence analogous to that found in human collagen. Amino acid analysis of purified CM4 protein express in *E. coli* strain JM109 (F-) under hydroxyproline-incorporating conditions compared to analysis of the same protein expressed under proline-incorporating conditions, demonstrates that the techniques herein result in essentially complete substitution of hydroxyproline for proline. The amino acid analysis was performed on CM4 protein that had been cleaved from and purified away from GST. This removes any possible ambiguities associated with the fusion protein.

[0180] Expression in media containing at least about 200 mM NaCl is preferable to accumulate significant amount of protein containing hydroxyproline. A concentration of about 400-500 mM Nacl appears to be optimal. Either KCl, sucrose or combinations thereof may be used in substitution of or with NaCl. However, expression in media without an added osmolyte (i.e. under conditions that more closely mimic those of Deming et al., In Vivo Incorporation of Proline Analogs into Artificial Protein, Poly. Mater. Sci. Engin. Proceed., supra.) did not result in significant expression of hydroxyproline-containing proteins in JM109 (F-). This is illustrated in Figure 36 which is a scan of a SDS-PAGE get showing the expression of GST-CM4 in media with or without 500 mM Nacl and containing either proline or hydroxy-

5

10

15

20

25

35

45

55

proline. The SDS-PAGE gel reflects 5 hour post-induction samples of GST-CM4 expressed in JM109 (F-). Equivalent amounts, based on OD600nm, of each culture were loaded in each lane. Gels were stained with Coomasie Blue, destained, and scanned on a PDI 420oe scanner. Lane 1: 2.5mM proline/0mM NaCl. Lane 2: 2.5mM proline/500mM NaCl. Lane 3: 80mM hydroxyproline/0mM NaCl. Lane 4: 80mM hydroxyproline/500mM NaCl. Lane 5: Molecular weight markers. The lower arrow indicates the migration position of proline-containing GST-CM4 in lanes 1 and 2. The upper arrow indicates the migration position of hydroxyproline-containing GST-CM4 in lanes 3 and 4. Note that GST-CM4 expressed in the presence of hydroxyproline runs at a higher apparent molecular weight (compare lanes 1 and 4). This is expected since hydroxyproline is of greater molecular weight than proline. If all the prolines in GST-CM4 are substituted with hydroxyproline, the increase in molecular weight is 671 Da (+2%). Note also that protein expressed in the presence of proline accumulates in cultures irrespective of the NaCl concentration (compare lanes 1 and 2). In contrast, significant expression in the presence of hydroxyproline only occurs in the culture containing 500 mM NaCl (compare lanes 3 and 4). Figure 37 further illustrates the dependence of expression on Nacl concentration by showing that significant expression of GST-CM4 occurs only at Nacl concentration greater than 200 mM. The SDS-PAGE gel reflects 6 hour post-induction samples of GST-CM4 expressed in JM109 (F-) with varying concentrations of NaCl. All cultures contained 80 mM hydroxyproline. Lane 1: 500 mM NaCl, not induced. Lanes 2-6: 500 mM, 400 mM, 300 mM, 200 mM, and 100 mM NaCl, respectively. All induced with 1.5 mM IPTG. Lane 7: Molecular weight markers. The arrow indicates the migration position of hydroxyproline-containing GST-CM4. Figure 38 is a scan of an SDS-PAGE gel of expression of GST-CM4 in either 400 mM NaCl or 800 mM sucrose. The SDS-PAGE gel reflects 4 hour post-induction samples of GST-CM4 expressed in JM109 (F-). All cultures contained 80 mM hydroxyproline and all, except that electrophoresed in lane 2, contained 400 mM NaCl. Lane 2 demonstrates expression in sucrose in lieu of NaCl. Lane 1: Molecular weight markers. Lane 2: 800 mM sucrose (no NaCl). Lanes 3-9: 0 mM, 0.025 mM, 0.1 mM, 0.4 mM, 0.8 mM, 1.25 mM, 2.5 mM proline, respectively. The upper arrow indicates the migration position of hydroxyproline-containing GST-CM4 and the lower arrow indicates the migration position of proline-containing GST-CM4. Expression is apparent in both cases (compare lanes 2 and 3).

[0181] If expression of GST-CM4, as described in Example 17 below, is performed in varying ratios of hydroxyproline and proline the expressed protein appears to contain varying amounts of hydroxyproline. Thus, if only hydroxyproline is present during expression, a single expressed protein of the expected molecular weight is evident on a SDS-PAGE gel (Figure 38, lane 3). If greater than approximately 1 mM proline is present, again a single expressed protein is evident, but at a lower apparent molecular weight, as expected for the protein containing only proline (Figure 38, lanes 7-9). If lesser amount of proline are used during expression, species of apparent molecular weight intermediate between these extremes are evident. This phenomenon, evident as a "smear" or "ladder" of proteins running between the two molecular weight extremes on an SDS-PAGE gel, is illustrated in lanes 3-9 of Figure 38. Lanes 3-9 on this gel are proteins from expression in a fixed concentration of 80 mM hydroxyproline and 400 mM NaCl. However, in moving from lane 3 to 9 the proline concentration increases from none (lane 3) to 2.5 mM (lane 9) and expression shifts from a protein of higher molecular weight (hydroxyproline-containing GST-CM4) to lower molecular weight (proline-containing GST-CM4). At proline concentrations of 0.025 mM and 0.1 mM, species of intermediate molecular weight are apparent (lanes 4 and 5). This clearly demonstrates that the percent incorporation of hydroxyproline in an expressed protein can be controlled by expression in varying ratios of analogue to amino acid.

[0182] Proline starvation prior to hydroxyproline incorporation is an important technique used herein. It insures that no residual proline is present during expression to compete with hydroxyproline. This enables essentially 100% substitution with the analogue. As shown in Figure 38, starvation conditions allow expression under precisely controlled ratios of proline and hydroxyproline. The amount of hydroxyproline vs. proline incorporated into the recombinant protein can therefore be controlled. Thus, particular properties of the recombinant protein that depend upon the relative amount of analogue incorporated can be tailored by the present methodology to produce polypeptides with unique and beneficial properties.

[0183] Human collagen, collagen fragments, collagen-like peptides (collagen mimetics) and the above chimeric polypeptides produced by recombinant processes have distinct advantages over collagen and its derivatives obtained from non-human animals. Since the human gene is used, the collagen will not act as a xenograft in the context of a medical implant. Moreover, unlike naturally occurring collagen, the extent of proline hydroxylation can be predetermined. This unprecedented degree of control permits detailed investigation of the contribution of *trans*-4-hydroxyproline to triple helix stabilization, fibril formation and biological activity. In addition, design of medical implants based upon the desired strength of collagen fibrils is enabled.

[0184] The following examples are included for purposes of illustration and are not to be construed as limitations herein.

EXAMPLE 1

Trans-membrane Transport

[0185] A 5 mL culture of E. coli strain DH5α (supE44 ΔlacU169 (φ80lacZ ΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1) containing a plasmid conferring resistance to ampicillin (pMAL-c2, Fig. 1) was grown in Luria Broth to confluency (~16 hours from inoculation). These cells were used to inoculate a 1 L shaker flask containing 500 mL of M9 minimal medium (M9 salts, 2% glucose, 0.01 mg/mL thiamine, 100 µg/mL ampicillin supplemented with all amino acids at 20 µg/mL) which was grown to an AU₆₀₀ of 1.0 (18-20 hours). The culture was divided in half and the cells harvested by centrifugation. The cells from one culture, were resuspended in 250 mL M9 media and those from the other in 250 mL of M9 media containing 0.5M NaCl. The cultures were equilibrated in an air shaker for 20 minutes at 37 °C (225 rpm) and divided into ten 25 mL aliquots. The cultures were returned to the shaker and 125 µl of 1M hydroxyproline in distilled H₂0 was added to each tube. At 2, 4, 8, 12, and 20 minutes, 4 culture tubes (2 isotonic, 2 hypertonic) were vacuum filtered onto 1 µm polycarbonate filters that were immediately placed into 2 mL microfuge tubes containing 1.2 mL of 0.2M NaOH/2% SDS in distilled H₂0. After overnight lysis, the filters were carefully removed from the tubes, and the supernatant buffer was assayed for hydroxyproline according to the method of Grant, Journal of Clinical Pathology, 17:685 (1964). The intracellular concentration of trans-4-hydroxyproline versus time is illustrated graphically in Figure 2.

20 EXAMPLE 2

25

30

35

45.

50

Effects of Salt Concentration on Transmembrane Transport

[0186] To determine the effects of salt concentration on transmembrane transport, an approach similar to Example 1 was taken. A 5 mL culture of *coli* strain DH5 α (supE44 Δlac U169 (ϕ 80/lacZ Δ M15) hsdR17 recA1 ental gyrA96 thi-1 relA1) containing a plasmid conferring resistance to ampicillin (pMAL-c2, Fig. 1) was grown in Luria Broth to confluency (~16 hours from inoculation). These cells were used to inoculate a 1 L shaker flask containing 500 mL of M9 minimal medium (M9 salts, 2% glucose, 0.01 mg/mL thiamine, 100 μ g/mL ampicillin supplemented with all amino acids at 20 μ g/mL) that was then grown to an AU $_{600}$ of 0.6. The culture was divided into three equal parts, the cells in each collected by centrifugation and resuspended in 150 mL M9 media, 150 mL M9 media containing 0.5M NaCl, and 150 mL M9 media containing 1.0M NaCl, respectively. The cultures were equilibrated for 20 minutes on a shaker at 37° C (225rpm) and then divided into six 25 mL aliquots. The cultures were returned to the shaker and 125 μ L of 1M hydroxyproline in distilled H $_2$ 0 was added to each tube. At 5 and 15 minutes, 9 culture tubes (3 isotonic, 3 x 0.5M NaCl, and 3 x 1.0M NaCl) were vacuum filtered onto 1 μ m polycarbonate filters that were immediately placed into 2 mL microfuge tubes containing 1.2 mL of 0.2M NaOH/2% SDS in distilled H $_2$ 0. After overnight lysis, the filters were removed from the tubes and the supernatant buffer assayed for hydroxyproline according to the method of Grant, supra.

EXAMPLE 2A

40 Effects of Salt Concentration on Transmembrane Transport

[0187] To determine the effects of salt concentration on transmembrane transport, an approach similar to Example 1 was taken. A saturated culture of JM109 (F-) harboring plasmid pD4 (Fig. 48) growing in Luria Broth (LB) containing 100μg/ml ampicillin (Amp) was used to inoculate 20 ml cultures of LB/Amp to an OD at 600 nm of 0.1 AU. The cultures were grown with shaking at 37°C to an OD 600 nm between 0.7 and 1.0 AU. Cells were collected by centrifugation and washed with 10 ml of M9 media. Each cell pellet was resuspended in 20 ml of M9/Amp media supplemented with 0.5% glucose and 100μg/ml of all of the amino acids except proline. Cultures were grown at 37°C for 30 min. to deplete endogenous proline. After out-growth, Nacl was added to the indicated concentration, Hyp was added to 40mM, and IPTG to 1.5mM. After 3 hours at 37°C, cells from three 5 ml aliquots of each culture were collected separately on polycarbonate filters and washed twice with five ml of M9 media containing 0.5% glucose and the appropriate concentration of NaCl. Cells were lysed in 1 ml of 70% ethanol by vortexing for 30 min. at room temperature. Cell lysis supernatants were taken to dryness, resuspended in 100μl of 2.5 N NaOH, and assayed for Hyp by the method of Neuman and Logan, R.E. Neuman and M.A. Logan, Journal of Biological Chemistry, 184:299 (1950). Total protein was determined with the BCA kit (Pierce, Rockford II) after cell lysis by three sonication/freeze-thaw cycles. The data are the mean ± standard error of three separate experiments. The intracellular concentration of *trans*-4-hydroxyproline versus NaCl concentration is illustrated graphically in Figure 2A.

EXAMPLE 3

5

10

15

20

25

30

35

45

50

55

Determination Of Proline Starvation Conditions in E. Coli

[0188] Proline auxotrophic *E. coli* strain NM519 (*pro*-) including plasmid pMAL-c2 which confers ampicillin resistance was grown in M9 minimal medium (M9 salts, 2% glucose, 0.01 mg/mL thiamine, 100μg mL ampicillin supplemented with all amino acids at 20 μg/mL except proline which was supplemented at 12.5 mg/L) to a constant AU₆₀₀ of 0.53 AU (17 hours post-inoculation). Hydroxyproline was added to 0.08<u>M</u> and hydroxyproline-dependent growth was demonstrated by the increase in the OD₆₀₀ to 0.61 AU over a one hour period.

EXAMPLE 4

Hydroxyproline Incorporation Into Protein in E. coli Under Proline Starvation Conditions

[0189] Plasmid pMAL-c2 (commercially available from New England Biolabs) containing DNA encoding for maltose-binding protein (MBP) was used to transform proline auxotrophic *E. coli* strain NM519 (*pro*-). Two 1 L cultures of transformed NM519 (*pro*-) in M9 minimal medium (M9 salts, 2% glucose, 0.01 mg/mL thiamine, 100 μg/mL ampicillin supplemented with all amino acids at 20 μg/mL except proline which was supplemented at 12.5 mg/L) were grown to an AU₆₀₀ Of 0.53 (~17 hours post-inoculation). The cells were harvested by centrifugation, the media in one culture was replaced with an equal volume of M9 media containing 0.08M hydroxyproline and the media in the second culture was replaced with an equal volume of M9 media containing 0.08M hydroxyproline and 0.5M NaCl. After a one hour equilibration, the cultures were induced with 1mM isopropyl-β-D-thiogalactopyranoside. After growing for an additional 3.25 hours, cells were harvested by centrifugation, resuspended in 10 mL of 10mM Tris-HCl (pH 8), 1mM EDTA, 100mM NaCl (TEN buffer), and lysed by freezing and sonication. MBP was purified by passing the lysates over 4 mL amylose resin spin columns, washing the columns with 10 mL of TEN buffer, followed by elution of bound MBP with 2 mL of TEN buffer containing 10mM maltose. Eluted samples were sealed in ampules under nitrogen with an equal volume of concentrated HCl (11.7M) and hydrolysed for 12 hours at 120 °C. After clarification with activated charcoal, hydroxyproline content in the samples was determined by HPLC and the method of Grant, *supra*. The percent incorporation of *trans*-4-hydroxyproline compared to proline into MBP is shown graphically in Figure 12.

EXAMPLE 5

Hydroxyproline Incorporation Into Protein in S. cerevisiae via Integrating Vectors Under Proline Starvation Conditions

[0190] The procedure described in Example 4 above is performed in yeast using an integrating vector which disrupts the proline biosynthetic pathway. A gene encoding human Type $1(\alpha_1)$ collagen is inserted into a unique shuttle vector behind the inducible GAL10 promoter. This promoter/gene cassette is flanked by a 5' and 3' terminal sequence derived from a S. cerevisiae proline synthetase gene. The plasmid is linearized by restriction digestion in both the 5' and 3' terminal regions and used to transform a proline-prototrophic S. cerevisiae strain. The transformation mixture is plated onto selectable media and transformants are selected. By homologous recombination and gene disruption, the construct simultaneously forms a stable integration and converts the S. cerevisiae strain into a proline auxotroph. A single transformant is selected and grown at 30 °C in YPD media to an OD_{600} of 2 AU. The culture is centrifuged and the cells resuspended in yeast dropout media supplemented with all amino acids except proline and grown to a constant OD_{600} indicating proline starvation conditions. 0.08M L-hydroxyproline and 2% (w/v) galactose is then added. Cultures are grown for an additional 6-48 hours. Cells are harvested by centrifugation (5000 rpm, 10 minutes) and lysed by mechanical disruption. Hydroxyproline-containing human Type $1(\alpha_1)$ collagen is purified by ammonium sulfate fractionation and column chromatography.

EXAMPLE 6

Hydroxyproline Incorporation Into Protein in *S. cerevisiae* via Non-Integrating Vectors Under Proline Starvation Conditions

[0191] The procedure described above in Example 4 is performed in a yeast proline auxotroph using a non-integrating vector. A gene encoding human Type 1 (α_1) collagen is inserted behind the inducible GAL10 promoter in the YEp24 shuttle vector that contains the selectable Ura* marker. The resulting plasmid is transformed into proline auxotrophic S. cerevisiae by spheroplast transformation. The transformation mixture is plated on selectable media and transformants are selected. A single transformant is grown at 30 °C in YPD media to an OD600 of 2 AU. The culture is centrifuged

and the cells resuspended in yeast dropout media supplemented with all amino acids except proline and grown to a constant OD_{600} indicating proline starvation conditions. 0.08M L-hydroxyproline and 2% (w/v) galactose is then added. Cultures are grown for an additional 6-48 hours. Cells are harvested by centrifugation (5000 rpm, 10 minutes) and lysed by mechanical disruption. Hydroxyproline-containing human Type 1 (α_1) collagen is purified by ammonium sulfate fractionation and column chromatography.

EXAMPLE 7

5

10

15

20

Hydroxyproline Incorporation Into Protein in a Baculovirus Expression System

[0192] A gene encoding human Type $1(\alpha_1)$ collagen is inserted into the pBacPAK8 baculovirus expression vector behind the AcMNPV polyhedron promoter. This construct is co-transfected into SF9 cells along with linearized AcMNPV DNA by standard calcium phosphate co-precipitation. Transfectants are cultured for 4 days at 27 °C in TNM-FH media supplemented with 10 % FBS. The media is harvested and recombinant virus particles are isolated by a plaque assay. Recombinant virus is used to infect 1 liter of SF9 cells growing in Grace's media minus proline supplemented with 10% FBS and 0.08 M hydroxyproline. After growth at 27 °C for 2-10 days, cells are harvested by centrifugation and lysed by mechanical disruption.

Hydroxyproline-containing human Type 1 (α_1) collagen is purified by ammonium sulfate fractionation and column chromatography.

EXAMPLE 8

Hydroxyproline Incorporation Into Human Collagen Protein in Escherichia coli Under Proline Starvation Conditions

25 [0193] A plasmid (pHuCol, Fig. 4) encoding the gene sequence of human Type I (α₁) collagen (Figures 3A and 3B) (SEQ. ID. NO. 1) placed behind the isopropyl-β-D-thiogalactopyranoside (IPTG)-inducible tac promotor and also encoding β-lactamase is transformed into Escherichia coli proline auxotrophic strain NM519 (pro-) by standard heat shock transformation. Transformation cultures are plated on Luria Broth (LB) containing 100 µg/ml ampicillin and after overnight growth a single ampicillin-resistant colony is used to inoculate 5 ml of LB containing 100 µg/ml ampicillin. After 30 growth for 10-16 hours with shaking (225 rpm) at 37 °C, this culture is used to inoculate 1 L of M9 minimal medium (M9 salts, 2% glucose, 0.01 mg/mL thiamine, 100 μg/mL ampicillin, supplemented with all amino acids at 20 μg/mL except proline which is supplemented at 12.5 mg/L) in a 1.5 L shaker flask. After growth at 37 °C, 225 rpm, for 15-20 hours post-inoculation, the optical density at 600 nm is constant at approximately 0.5 OD/mL. The cells are harvested by centrifugation (5000 rpm, 5 minutes), the media decanted, and the cells resuspended in 1 L of M9 minimal media containing 100 µg/mL ampicillin, 0.08M L-hydroxyproline, and 0.5M NaCl. Following growth for 1 hour at 37 °C, 225 rpm, IPTG is added to 1mM and the cultures allowed to grow for an additional 5-15 hours. Cells are harvested by centrifugation (5000 rpm, 10 minutes) and lysed by mechanical disruption. Hydroxyproline-containing collagen is purified by ammonium sulfate fractionation and column chromatography.

40 EXAMPLE 9

Hydroxyproline Incorporation Into Fragments of Human Collagen Protein in Escherichia coli Under Proline Starvation Conditions

45 [0194] A plasmid (pHuCol-FI, Figure 6) encoding the gene sequence of the first 80 amino acids of human Type 1 (α_1) collagen (Figure 5) (SEQ. ID. NO. 2) placed behind the isopropyl- β -D-thiogalactopyranoside (IPTG)-inducible tac promotor and also encoding β-lactamase is transformed into Escherichia coli proline auxotrophic strain NM519 (pro·) by standard heat shock transformation. Transformation cultures are plated on Luria Broth (LB) containing 100 μg/mL ampicillin and after overnight growth a single ampicillin-resistant colony is used to inoculate 5 mL of LB containing 100 50 μg/mL ampicillin. After growth for 10-16 hours with shaking (225 rpm) at 37 °C, this culture is used to inoculate 1 L of M9 minimal medium (M9 salts, 2% glucose, 0.01 mg/mL thiamine, 100 μg/mL ampicillin, supplemented with all amino acids at 20 µg/mL except proline which is supplemented at 12.5 mg/L) in a 1.5 L shaker flask. After growth at 37 °C, 225 rpm, for 15-20 hours post-inoculation, the optical density at 600 nm is constant at approximately 0.5 OD/mL. The cells are harvested by centrifugation (5000 rpm, 5 minutes), the media decanted, and the cells resuspended in 1 L of 55. M9 minimal media containing 100 μg/mL ampicillin, 0.08M L-hydroxyproline, and 0.5M NaCl. Following growth for 1 hour at 37 °C, 225 rpm, IPTG is added to 1mM and the cultures allowed to grow for an additional 5-15 hours. Cells are harvested by centrifugation (5000 rpm, 10 minutes) and lysed by mechanical disruption. The hydroxyproline-containing collagen fragment is purified by ammonium sulfate fractionation and column chromatography.

EXAMPLE 10

Construction and Expression in E. coli of the Human Collagen Type 1(a1) Gene with Optimized E. coli Codon Usage

5 A. Construction of the gene:

[0195] The nucleotide sequence of the helical region of human collagen Type I (α_1) gene flanked by 17 amino acids of the amino terminal extra-helical and 26 amino acids of the C-terminal extra-helical region is shown in Figure 27 (SEQ. ID. NO. 15). A tabulation of the codon frequency of this gene is given in Table I. The gene sequence shown in Figure 27 was first changed to reflect *E. coli* codon bias. An initiating methionine was inserted at the 5' end of the gene and a TAAT stop sequence at the 3' end. Unique restriction sites were identified or created approximately every 150 base pairs. The resulting gene (HUCol^{EC}, Figure 39A-39E) (SEQ. ID. NO. 20) has the codon usage given in Table II as shown below. Other sequences that approximate *E. coli* codon bias are also acceptable.

TABLE II

	Codon	Count	%age	Codon	Count	%age	Codon	Count	%age	Codon	Count	%age
	TTT-	6	0.56	тст-	3	0.28	TAT-	2	0.18	TGT-	0	0.00
	Phe			Ser			Tyr			Cys		
	TTC-	9	0.85	TCC-	3	0.28	TAC-	2	0.18	TGC-	0	0.00
	Phe			Ser			Tyr			Cys		
	TTA-	0	0.00	TCA-	0	0.00	TAA-	0	0.00	TGA-***	0	0.00
	Leu			Ser	·		***					
	TTG-	0	0.00	TCG-	0	0.00	TAG-	0	0.00	TGG-	0	0.00
	Leu			Ser			***			Trp		
	CTT-	0	0.00	CCT-	13	1.22	CAT-	0	0.00	CGT-	26	2.45
	Leu			Pro			His			Arg		
	CTC-	1	0.09	CCC-	12	1.13	CAC-	3	0.28	CGC-	26	2.45
)	Leu			Pro			His			Arg		
	CTA-	1	0.09	CCA-	29	2.74	CAA-	5	0.47	CGA-	0	0.00
	Leu			Pro			Gln			Arg		
	CTG-	19	1.79	CCG-	186	17.58	CAG-	25	2.36	CGG-	1	0.09
	Leu			Pro			Gln			Arg		
i [;]	ATT-	3	0.28	ACT-	2	0.18	AAT-	0	0.00	AGT-	1	0.09
	lle			Thr			Asn			Ser		
	ATC-	4	0.37	ACC-	11	1.03	AAC-	11	1.03	AGC-	32	3.02
	lle			Thr			Asn			Ser		
۱۰	ATA-	0	0.00	ACA-	0	0.00	AAA-	38	3.59	AGA-	0	0.00
,	lle			Thr			Lys			Arg		
	ATG-	8	0.75	ACG-	4	0.37	AAG-	0	0.00	AGG-	0	0.00
	Met			Thr			Lys			Arg		
	GTT-	3	0.28	GCT-	10	0.94	GAT-	20	1.89	GGT-	148	13.98
i	Val			Ala			Asp			Gly		
	GTC-	5	0.47	GCC-	24	2.26	GAC-	14	1.32	GGC-	178	16.82
	Val			Ala			Asp			Gly		
	GTA-	0	0.00	GCA-	8	0.75	GAA-	40	3.78	CGA-	9	0.85
,	Val			Ala			Glu	_		Gly		
,	GTG-	12	1.13	GCG-	80	7.56	GAG-	9	0.85	GGG-	12	1.13
	Val			Ala			Glu			Gly		

[0196] Oligos of approximately 80 nucleotides were synthesized on a Beckman Oligo 1000 DNA synthesizer, cleaved and deprotected with aqueous NH_4OH , and purified by electrophoresis in 7M urea/12% polyacrylamide gels. Each set of oligos was designed to have an EcoR I restriction enzyme site at the 5' end, a unique restriction site near the 3' end, followed by the TAAT stop sequence and a Hind III restriction enzyme site at the very 3' end. The first four oligos, comprising the first 81 amino acids of the human collagen Type I (α_1) gene, are given in Figure 40 which shows the

sequence and restriction maps of synthetic oligos used to construct the first 243 base pairs of the human Type I (α_1) collagen gene with optimized *E. coli* codon usage. Oligos N1-1 (SEQ. ID. NO. 21) and N1-2 (SEQ. ID. NO. 22) were designed to insert an initiating methionine (ATG) codon at the 5' end of the gene.

[0197] In one instance, oligos N1-1 and N1-2 (1µg each) were annealed in 20 µL of T7 DNA polymerase buffer (40mM Tris·HC1 (pH 8.0), 5mM MgCl₂, 5mM dithiothreitol, 50mM NaCl, 0.05 mg/mL bovine serum albumin) by heating at 90°C for 5 minutes followed by slow cooling to room temperature. After brief centrifugation at 14,000 rpm, 10 units of T7 DNA polymerase and 2 µL of a solution of all four dNTPs (dATP, dGTP, dCTP, dTTP, 2.5mM each) were added to the annealed oligos. Extension reactions were incubated at 37°C for 30 minutes and then heated at 70°C for 10 minutes. After cooling to room temperature, Hind III buffer (5 pL of 10x concentration), 20 μL of H₂O, and 10 units of Hind III restriction enzyme were added and the tubes incubated at 37°C for 10 hours. Hind III buffer (2µL of 10x concentration), 13.5μL of 0.5M Tris·HC1 (pH 7.5), 1.8 μL of 1% Triton X100, 5.6 μL of H₂O, and 20 U of EcoR I were added to each tube and incubation continued for 2 hours at 37°C. Digests were extracted once with an equal volume of phenol, once with phenol/chloroform/isoamyl alcohol, and once with chloroform/isoamyl alcohol. After ethanol precipitation, the pellet was resuspended in 10 µL of TE buffer (10mM Tris·HC1 (pH 8.0), 1mM EDTA). Resuspended pellet (4 µL) was ligated overnight at 16°C with agarose gel-purified EcoRI/Hind III digested pBSKS* vector (1μg) using T4 DNA ligase (100 units). One half of the transformation mixture was transformed by heat shock into DH5α cells and 100 μL of the 1.0 mL transformation mixture was plated on Luria Broth (LB) agar plates containing 70 µg/mL ampicillin. Plates were incubated overnight at 37°C. Ampicillin resistant colonies (6-12) were picked and grown overnight in LB media containing 70 mg/mL ampicillin. Plasmid DNA was isolated from each culture by Wizard Minipreps (Promega Corporation, Madison WI) and screened for the presence of the approximately 120 base pair insert by digestion with EcoR I and Hind III and running the digestion products on agarose electrophoresis gels. Clones with inserts were confirmed by standard dideoxy termination DNA sequencing. The correct clone was named pBSN1-1 (Figure 41) and the collagen fragment has the nucleic acid sequence given in Figure 42 (SEQ. ID. NO. 25).

[0198] Oligos N1-3 (SEQ. ID. NO. 23) and N1-4 (SEQ. ID. NO. 24) (Figure 40) were synthesized, purified, annealed, extended, and cloned into pBSKS⁺ following the same procedure given above for oligos N1-1 and N1-2. The resulting plasmid was named pBSN1-2A. To clone together the sections of the collagen gene from pBSN1-1 and pBSN1-2A, plasmid pBSN1-1 (1 μg) was digested for 2 hours at 37°C with Rsr II and Hind III. The digested vector was purified by agarose gel electrophoresis. Plasmid pBSN1-2A (3 μg) was digested for 2 hours at 37°C with Rsr II and Hind III and the insert purified by agarose gel electrophoresis. Rsr II/Hind III-digested pBSN1-1 was ligated with this insert overnight at 16°C with T4 DNA ligase. One half of the ligation mixture was transformed into DH5α cells and 1/10 of the transformation mixture was plated on LB agar plates containing 70 μg/mL ampicillin. After overnight incubation at 37°C, ampicillin-resistant clones were picked and screened for the presence of insert DNA as described above. Clones were confirmed by dideoxy termination sequencing. The correct clone was named pBSN1-2 (Figure 43) and the collagen fragment has the sequence given in Figure 44.

³⁵ [0199] In similar manner, the remainder of the collagen gene is constructed such that the final DNA sequence is that given in Figure 39A-39E (SEQ. ID. NO. 19).

B) Expression of the gene in E. coli:

[0200] Following construction of the entire human collagen Type I (α₁) gene with codon usage optimized for *E. coli*, the cloned gene is expressed in *E. coli*. A plasmid (pHuCol^{Ec}, Figure 45) encoding the entire synthetic'collagen gene (Figure 39A-39E) placed behind the isopropyl-β-D-thiogalactopyranoside (IPTG)-inducible *tac* promotor and also encoding β-lactamase is transformed into *Escherichia coli* strain DH5α (*supE44* Δ/acU169 (φ80/acZ ΔM15) *hsd*R17 *rec*A1 endA1 gyrA96 *thi-*1 *rel*A1) by standard heat shock transformation. Transformation cultures are plated on Luria Broth (LB) containing 100 μg/mL ampicillin and after overnight growth a single ampicillin-resistant colony is used to inoculate 10 mL of LB containing 100 μg/mL ampicillin. After growth for 10-16 hours with shaking (225 rpm) at 37°C, this culture is used to inoculate 1 L of LB containing 100 μg/mL ampicillin in a 1.5 L shaker flask. After growth at 37°C, 225 rpm, for 2 hours post-inoculation, the optical density at 600 nm is approximately 0.5 OD/mL. IPTG is added to 1mM and the culture allowed to grow for an additional 5-10 hours. Cells are harvested by centrifugation (5000 rpm, 10 minutes) and lysed by mechanical disruption. Recombinant human collagen is purified by ammonium sulfate fractionation and column chromatography. The yield is typically 15-25 mg/L of culture.

EXAMPLE 11

. 5

10

15

20

5 Expression in E. coli of an 81 Amino Acid Fragment of Human Collagen Type I(α1) with Optimized E. coli Codon Usage

[0201] A plasmid (pTrcN1-2, Figure 46) encoding the gene sequence of the first 81 amino acids of human Type I (α_1) collagen with optimized *E. coli* codon usage cloned in fusion with a 6 histidine tag at the 5' end of the gene and

placed behind the isopropyl- β -D-thiogalactopyranoside (IPTG)-inducible trc promotor and also encoding β -lactarnase was constructed by subcloning the EcoR I/Hind III insert from pBSN1-2 into the EcoR I/Hind III site of plasmid pTrcB (Invitrogen, San Diego, CA). Plasmid pTrcN1-2 was transformed into Escherichia coli strain DH5α (supE44ΔlacU169 (φ80/ac/Z ΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1) by standard heat shock transformation. Transformation cultures were plated on Luria Broth (LB) containing 100 µg/mL ampicillin and after overnight growth a single ampicillinresistant colony was used to inoculate 5mL of LB containing 100 µg/mL ampicillin. After growth for 10-16 hours with shaking (225 rpm) at 37°C, this culture was used to inoculate 50 mL of LB containing 100 μg/mL ampicillin in a 250 mL shaker flask. After growth at 37°C, 225 rpm, for 2 hours post-inoculation, the optical density at 600 nm was approximately 0.5 OD/mL. IPTG was added to 1mM and the culture allowed to grow for an additional 5-10 hours. Cells were harvested by centrifugation (5000 rpm, 10 minutes) and stored at -20°C. The 6 histidine tag-collagen fragment fusion was purified on nickel resin columns. Cell pellets were resuspended in 10 mL of 6M guanidine hydrochloride/ 20mM sodium phosphate/500mM Nacl (pH 7.8) and bound in two 5 mL batches to the nickel resin. Columns were washed two times with 4 mL of binding buffer (8M urea/20mM sodium phosphate/500mM NaCl (pH 7.8)), two times with wash buffer 1 (8M urea/20mM sodium phosphate/500mM NaCl (pH 6.0)), and two times with wash buffer 2 (8m urea/20mM sodium phosphate/500mM NaCl (pH 5.3). The 6 histidine tag-collagen fragment fusion was eluted from the column with 5mL of elution buffer (8M urea/20mM sodium phosphate/500mM NaCl (pH 4.0) in 1 mL fractions. Fractions were assessed for protein by gel electrophoresis and fusion-containing fractions were concentrated and stored at -20°C. The yield was typically 15-25 mg/L of culture.

[0202] The collagen is cleaved from the 6 histidine tag with enterokinase. Fusion-containing fractions are dialyzed against cleavage buffer (50mM Tris·HCI, pH 8.0/5mM CaCl₂). After addition of enterokinase at 1 μ g enzyme for each 100 μ g fusion, the solution is incubated at 37°C for 4-10 hours. Progress of the cleavage is monitored by gel electrophoresis. The cleaved 6 histidine tag may be separated from the collagen fragment by passage over a nickel resin column as outlined above.

EXAMPLE 12

., 5

20

25

30

40

Expression in E. coli of Fragments of Human Collagen Type I (α₁) with Optimized E. coli Codon Usage

[0203] A plasmid (pN1-3, Figure 47) encoding the gene for the amino terminal 120 amino acids of human collagen Type I (α₁) with optimized *E. coli* codon usage placed behind the isopropyl-β-D-thiogalactopyranoside (IPTG)-inducible *tac* promotor and also encoding β-lactamase is transformed into *Escherichia coli* strain DH5α (sup E44 ΔlacU169 (φ80/acZ ΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1) by standard heat shock transformation. Transformation cultures are plated on Luria Broth (LB) containing 100 μg/mL ampicillin and after overnight growth a single ampicillin-resistant colony is used to inoculate 10 mL of LB containing 100 μg/mL ampicillin. After growth for 10-16 hours with shaking (225 rpm) at 37°C, this culture is used to inoculate 1 L of LB containing 100 μg/mL ampicillin in a 1.5 L shaker flask. After growth at 37°C, 225 rpm, for 2 hours post-inoculation, the optical density at 600 nm is approximately 0.5 OD/mL. IPTG is added to 1mM and the culture allowed to grow for an additional 5-10 hours. Cells are harvested by centrifugation (5000 rpm, 10 minutes) and lysed by mechanical disruption. Recombinant human collagen is purified by ammonium sulfate fractionation and column chromatography. The yield is typically 15-25 mg/L of culture.

EXAMPLE 13

Expression in E. coli of a C-terminal Fragment of Human Collagen Type I (α_1) with Optimized E. coli Codon Usage.

[0204] A plasmid (pD4, Figure 48) encoding the gene for the carboxy terminal 219 amino acids of human collagen Type I (α₁) with optimized *E. coli* codon usage placed behind the isopropyl-β-D-thiogalactopyranoside (IPTG)-inducible tac promotor and also encoding β-lactamase is transformed into *Escherichia coli* strain DH5α (sup E44 Δ/acU169 (φ80/acZ ΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1) by standard heat shock transformation. Transformation cultures are plated on Luria Broth (LB) containing 100 μg/mL ampicillin and after overnight growth a single ampicillin-resistant colony is used to inoculate 10 mL of LB containing 100 μg/mL ampicillin. After growth for 10-16 hours with shaking (225 rpm) at 37°C, this culture is used to inoculate 1 L of LB containing 100 μg/mL ampicillin in a 1.5 L shaker flask. After growth at 37°C, 225 rpm, for 2 hours post-inoculation, the optical density at 600 nm is approximately 0.5 OD/mL. IPTG is added to 1mM and the culture allowed to grow for an additional 5-10 hours. Cells are harvested by centrifugation (5000 rmp, 10 minutes) and lysed by mechanical disruption. Recombinant human collagen fragment is purified by ammonium sulfate fractionation and column chromatography. The yield is typically 15-25 mg/L of culture.

EXAMPLE 14

Construction and Expression in E. coli of the Human Collagen Type 1 (a2) Gene with Optimized E. coli Codon Usage

A) Construction of the gene:

[0205] The nucleotide sequence of the helical region of human collagen Type I (α_2) gene flanked by 11 amino acids of the amino terminal extra-helical and 12 amino acids of the C-terminal extra-helical region is shown in Figures 49A-49E (SEQ. ID. NO. 29). A tabulation of the codon frequency of this gene is given in Table III below. The gene sequence shown in Figures 49A-49E was first changed to reflect *E. coli* codon bias. An initiating methionine was inserted at the 5' end of the gene and a TAAT stop sequence at the 3' end. Unique restriction sites are identified or created approximately every 150 base pairs. The resulting gene (HuCol(α_2)^{Ec}, Figures 50A-50E) (SEQ. ID. NO. 31) has the codon usage given in Table IV below. Other sequences that approximate *E. coli* codon bias are also acceptable.

Table III

	Codon	Count	\$age	Codon	Count	\$age	Codon	Count	åage	Codon	Count	\$age
ļ	TTT-Phe	3	0.28	TCT-Ser	11	1.06	TAT-Tyr	2	0.19	TGT-Cys	0	0.00
	TTC-Phe	10	0.96	TCC-Ser	4	0.38	TAC-Tyr	3	0.28	TGC-Cys	0	0.00
	TTA-Leu	1	0.09	TCA-Ser	1	0.09	TAA-***	0	0.00	TGA-+++	0	0.00
	TTG-Leu	2	0.19	TCG-Ser	1	0.09	TAG-***	0	0.00	TGG-Trp	0	0.00
	CTT-Leu	16	1.54	CCT-Pro	125	12.06	CAT-His	7	0.67	CGT-Arg	17	1.64
1	CTC-Leu	9	0.86	CCC-Pro	42	4.05	CAC-His	6	0.57	CGC-Arg	6	0.57
	CTA-Leu	2	0.19	CCA-Pro	30	2.89	CAA-Gln	13	1.25	CGA-Arg	6	0.57
	CTG-Leu	5	0.48	CCG-Pro	3	0.28	CAG-Gln	9	0.86	CGG-Arg	4	0.38
	ATT-Ile	14	1.35	ACT-Thr	14	1.35	AAT-Asn	10	0.96	AGT-Ser	11	1.06
	ATC-Ile	3	0.28	ACC-Thr	0	0.00	AAC-Aan	14	1.35	AGC-Ser	4	0.38
- 1	ATA-Ile	1	0.09	ACA-Thr	3	0.28	AAA-Lys	15	1.44	AGA-Arg	16	1.54
-	ATG-Met	5	0.48	ACG-Thr	1	0.09	AAG-Lys	16	1.54	AGG-Arg	6	0.57
- 1	GTT-Val	20	1.93	GCT-Ala	82	7.91	GAT-Asp	20	1.93	GGT-Gly	179	17.27
- 1	GTC-Val	5	0.48	GCC-Ala	3.7	1.64	GAC-Asp	5	0.48	GGC-Gly	74	7.14
١	GTA-Val	3	0.28	GCA-Ala	9	0.86	GAA-Glu	29	2.79	GGA-Gly	80	7.72
ĺ	GTG-Val	10	0.96	GCG-Ala	0	0.00	GAG-Glu	16	1.54	GGG-Gly	16	1.54

.;

Table IV

Codon	Count	\$age	Codon	Count	हैबपुरु	Codon	Count	889e	Codon	Count	ваде
TTT-Phe	5	0.48	TCT-Ser	7	0.67	TAT-TYY	3	0.28	TGT-Cys	0	0.00
TTC-Phe	7	0.67	TCC-Ser	12	1.15	TAC-Tyr	2	0.19	TGC-Cys	0	0.00
TTA-Leu	0	0.00	TCA-Ser	0	0.00	TAA-***	0	0.00	TGA-***	0	0.00
TTG-Leu	0	0.00	TCG-Ser	0	0.00	TAG-***	0	0.00	TGG-Trp	0	0.00
CTT-Leu	1	0.09	CCT-Pro	10	0.96	CAT-His	2	0.19	CGT-Arg	37	3.55
CTC-Leu	1	0.09	CCC-Pro	0	0.00	CAC-His	11	1.05	CGC-Arg	18	1.72
CTA-Leu	0	0.00	CCA-Pro	15	1.44	CAA-Gln	7	0.67	CGA-Arg	0	0.00
CTG-Leu	32	3.07	CCG-Pro	177	17.00	CAG-Gln	15	1.44	CGG-Arg	0	0.00
ATT-Ile	11	1.05	ACT-Thr	3	0.28	AAT-Asn	6	0.57	AGT-Ser	Ō	0.00
ATC-Ile	7	0.67	ACC-Thr	6	0.57	AAC-Asn	18	1.72	AGC-Ser	13	1.24
ATA-Ile	0	0.00	ACA-Thr	0	0.00	AAA-Lys	25	2.40	AGA-Arg	0	0.00
ATG-Met	6	0.57	ACG-Thr	10	0.96	AAG-Lys	6	0.57	AGG-Arg	0	0.00
GTT-Val	18	1.72	GCT-Ala	30	2.88	GAT-Asp	11	1.05	GGT-Gly	209	20.07
GTC-Val	7	0.67	GCC-Ala	21	2.01	GAC-Asp	13	1.24	GGC-Gly	141	13.54
GTA-Val	9	0.85	GCA-Ala	20	1.92	GAA-Glu	33	3.17	GGA-Gly	0	0.00
GTG-Val	6	0.57	GCG-Ala	38	3.65	GAG-Glu	12	1.15	GGG-Gly	0	0.00

5

10.

15

20.

25

35

40

45

[0206] Oligos of approximately 80 nucleotides are synthesized on a Beckman Oligo 1000 DNA synthesizer, cleaved and deprotected with aqueous NH₄OH, and purified by electrophoresis in 7M urea/12% polyacrylamide gels. Each set of oligos is designed to have an EcoR I restriction enzyme site at the 5' end, a unique restriction site near the 3' end; followed by the TAAT stop sequence and a Hind III restriction enzyme site at the very 3' end. Oligos N1-1(α_2) and N1-2 (α_2) are designed to insert an initiating methionine (ATG) codon at the 5' end of the gene.

[0207] In one instance, oligos N1-1(α_2) and N1-2(α_2) (1 μg each) (Figure 51 depicts sequence and restriction maps of synthetic oligos used to construct the first 240 base pairs of human Type I(α2) collagen gene with optimized E. coli codon usage) are annealed in 20 µL of T7 DNA polymerase buffer (40mM Tris·HCI (pH 8.0), 5mM MgCl₂, 5mM dithiothreitol, 50mM NaCl, 0.05 mg/mL bovine serum albumin) by heating at 90°C for 5 minutes followed by slow cooling to room temperature. After brief centrifugation at 14,000 rpm, 10 units of T7 DNA polymerase and 2 μL of a solution of all four dNTPs (dATP, dGTP, dCTP, dTTP, 2.5mM each) are added to the annealed oligos. Extension reactions are incubated at 37°C for 30 minutes and then heated at 70°C for 10 minutes. After cooling to room temperature, Hind III buffer (5 μ L of 10x concentration), 20 μ L of H₂O, and 10 units of Hind III restriction enzyme are added and the tubes incubated at 37°C for 10-16 hours. Hind III buffer (2 μL of 10x concentration), 13.5 μL of 0.5 Tris·HCI (pH 7.5); 1.8 μL of 1% Triton X100, 5.6 µL of H₂O, and 20 U of EcoR I are added to each tube and incubation continued for 2 hours at 37°C. Digests are extracted once with an equal volume of phenol, once with phenol/chloroform/isoamyl alcohol, and once with chloroform/isoamyl alcohol. After ethanol precipitation, the pellet is resuspended in 10 µL of TE buffer (10mM Tris·HCI (pH 8.0), 1mM EDTA). Resuspended pellet (4 µL) is ligated overnight at 16°C with agarose gel-purified EcoRI/ Hind III digested pBSKS+ vector (1 µg) using T4 DNA ligase (100 units). One half of the transformation mixture is transformed by heat shock into DH5 α cells and 100 μ L of the 1.0 mL transformation mixture is plated on Luria Broth (LB) agar plates containing 70 μg/mL ampicillin. Plates are incubated overnight at 37°C. Ampicillin resistant colonies (6-12) are picked and grown overnight in LB media containing 70 μg/mL ampicillin. Plasmid DNA is isolated from each culture by Wizard Minipreps (Promega Corporation, Madison, WI) and screened for the presence of the approximately 120 base pair insert by digestion with EcoR I and Hind III and running the digestion products on agarose electrophoresis gels. Clones with inserts are confirmed by standard dideoxy termination DNA sequencing. The correct clone is named pBSN1-1(α_2) Figure 52).

[0208] Oligos N1-3(α_2) and N1-4(α_2) are synthesized, purified, annealed, extended, and cloned into pBSKS+ following the same procedure given above for oligos N1-1(α_2) and N1-2(α_2). The resulting plasmid is named pBSN1-2A. To clone together the sections of the collagen gene from pBSN1-1(α_2) (1 μ g) is digested for 2 hours at 37°C with BsrF I and Hind III. The digested vector is purified by agarose gel electrophoresis. Plasmid pBSn1-2(α_2) (3 μ g) is digested for 2 hours at 37°C with BsrF I and Hind III and the insert purified by agarose gel electrophoresis. BsrF I/Hind III-digested pBSN1-1 is ligated with this insert overnight at 16°C with T4 DNA ligase. One half of the ligation mixture is transformed into DH5 α cells and 1/10 of the transformation mixture is plated on LB agar plates containing 70 μ g/mL ampicillin. After overnight incubation at 37°C, ampicillin-resistant clones are picked and screened for the presence of insert DNA as described above. Clones are confirmed by dideoxy termination sequencing. The correct clone is name

pBSN1-2(α_2) (Figure 53) and the collagen fragment has the sequence given in Figure 54 (SEQ. ID. NO. 37). [0209] In a similar manner, the remainder of the collagen gene is constructed such that the final DNA sequence is that given in Figures 50A-50E (SEQ. ID. NO. 31).

B) Expression of the gene in E. coli:

[0210] Following construction of the entire human collagen Type I (α 2) gene with codon usage optimized for *E. coli*, the cloned gene is expressed in *E. coli*. A plasmid (pHucol(α_2)^{Ec}, Figure 55) encoding the entire synthetic collagen gene (Figures 50A-50E) placed behind the isopropyl- β -D-thiogalactopyranoside (IPTG)-inducible *tac* promotor and also encoding β -lactamase is transformed into *Escherichia coli* strain DH5 α (supE44 Δ IacU169 (ϕ 80/IacZ Δ M15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1) by standard heat shock transformation. Transformation cultures are plated on Luria Broth (LB) containing 100 μ g/mL ampicillin and after overnight growth a single ampicillin-resistant colony is used to inoculate 10 mL of LB containing 100 μ g/mL ampicillin. After growth for 10-16 hours with shaking (225 rpm) at 37°C, this culture is used to inoculate 1 L of LB containing 100 μ g/mL ampicillin in a 1.5 L shaker flask. After growth at 37°C, 225 rpm, for 2 hours post-inoculation, the optical density at 600 nm is approximately 0.5 OD/mL. IPTG is added to 1mM and the culture allowed to grow for an additional 5-10 hours. Cells are harvested by centrifugation (5000 rpm, 10 minutes) and lysed by mechanical disruption. Recombinant human collagen is purified by ammonium sulfate fractionation and column chromatography. The yield is typically 15-25 mg/L of culture.

EXAMPLE 14A

5

10

15

20

25

30

45

55

Alternative Construction and Expression in *E. Coli* of the Human Collagen Type 1 (α2) Gene with Optimized *E. coli* Codon Usage

A) Construction of the gene:

[0211] The nucleotide sequence of the helical region of human collagen Type 1 (α 2) gene flanked by 11 amino acids of the amino terminal extra-helical and 12 amino acids of the C-terminal extra-helical region is shown in Figures 49A-49E (SEQ. ID. NO. 29). A tabulation of the codon frequency of this gene is given in Table III. The gene sequence shown in Figures 49A-49E was first changed to reflect *E. coli* codon bias. An initiating methionine was inserted at the 5' end of the gene and a TAAT stop sequence at the 3' end. Unique restriction sites were identified or created at appropriate locations in the gene (approximately every 150 base pairs). The resulting gene (HuCol(α 2)^{EC}, Figures 50A-50E) (SEQ. ID. NO. 31) has the codon usage given in Table IV. Other sequences that approximate *E. coli* codon bias are also acceptable.

[0212] Oligonucleotides were synthesized on a Beckman Oligo 1000 DNA synthesizer, cleaved and deprotected with aqueous NH_AOH, and purified by electrophoresis in 7M urea/12% polyacrylamide gels. Purified oligos (32.5 pmol) were dissolved in 20µL of ligation buffer (Boehringer Mannheim, Cat. No. 1635 379) and annealed by heating to 95°C followed by slow cooling to 20°C over 45 minutes. The annealed oligonucleotides were ligated for 5 minutes at room temperature with digested vector (1µg) using T4 DNA ligase (5 units). One half of the transformation mixture was transformed by heat shock into DH5 α cells and 100 μ L of the 1.0mL transformation mixture plated on Luria Broth (LB) agar plates containing 70µg/mL ampicillin. Plates were incubated overnight at 37°C. Ampicillin resistant colonies (6-12) were picked and grown overnight in LB media containing 70µg/mL ampicillin. Plasmid DNA was isolated from each culture by QIAprep Miniprep (Qiagen, Valencia, CA) and screened for the presence of insert by digestion with flanking restriction enzymes and running the digestion products on agarose electrophoresis gels. Clones with inserts were confirmed by standard dideoxy termination DNA sequencing. To clone together the sections of the collagen gene, and insert covering a flanking portion of the gene was ligated into vector containing the neighboring gene portion. Inserts were isolated from plasmids and vectors were cut by double digestion for 2 hours at 37°C with the appropriate restriction enzymes. The digested vector and insert were purified by agarose gel electrophoresis. Insert and vector were ligated for 5 minutes at room temperature following the procedure in the Rapid DNA Ligation Kit (Boehringer Mannheim). One half of the ligation mixture is transformed into DH5α cells and 1/10 of the transformation mixture was plated on LB agar plates containing 70µg/mL ampicillin. After overnight incubation at 37°C, ampicillin-resistant clones were picked and screened for the presence of insert DNA as described above. Clones were confirmed by dideoxy termination sequencing.

[0213] In a similar manner, the remainder of the collagen gene was constructed such that the final DNA sequence is that given in Figures 50A-50E (SEQ. ID. NO. 31).

B) Expression of the gene in E. coli:

[0214] Following construction of the entire human collagen Type $1(\alpha 2)$ gene with codon usage optimized for *E. coli*, the cloned gene is expressed in *E. coli*. A plasmid (pHuCol)($\alpha 2$)^{Ec}, Figure 55) encoding the entire collagen gene (Figures 50A-50E) placed behind the isopropyl- β -D-thiogalactopyranoside (IPTG)-inducible *tac* promoter and also encoding β -lactamase is transformed into *Escherichia coli* strain DH5 α (sup*E*44 Δ /acU169 (ϕ 80/acZ Δ M15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1) by standard heat shock transformation. Transformation cultures are plated on Luria Broth (LB) containing 100 μ g/mL ampicillin and after overnight growth a single ampicillin-resistant colony is used to inoculate 10 mL of LB containing 100 μ g/mL ampicillin. After growth for 10-16 hours with shaking (225 rpm) at 37°C, this culture is used to inoculate 1 L of LB containing 100 μ g/mL ampicillin in a 1.5 L shaker flask. After growth at 37°C, 225 rpm, for 2 hours post-inoculation, the optical density at 600 nm is approximately 0.5 OD/mL. IPTG is added to 1mM and the culture allowed to grow for an additional 5-10 hours. Cells are harvested by centrifugation (5000 rpm, 10 minutes) and lysed by mechanical disruption. Recombinant human collagen is purified by ammonium sulfate fractionation and column chromatograph. The yield is typically 15-25 mg/L of culture.

EXAMPLE 15

5

10

15

20

25

30

ä

45

50

55

Expression in E. coli of Fragments of Human Collagen Type I(α2) with Optimized E. coli Codon Usage

[0215] A plasmid (pN1-2, Figure 56) encoding the gene for the amino terminal 80 amino acids of human collagen Type I(α₂) (SEQ. ID. NO. 31, Fig. 54) with optimized *E. coli* codon usage placed behind the isopropyl-β-D-thiogalact-opyranoside (IPTG)-inducible *tac* promotor and also encoding (β-lactamase is transformed into *Escherichia coli* strain DH5α (*sup*E44 Δ/*ac*U169 (φ80/*ac*Z ΔΜ15) *hsd*R17 *rec*A1 *end*A1 *gyr*A96 *thi*-1 *rel*A1) by standard heat shock transformation. Transformation cultures are plated on Luria Broth (LB) containing 100 μg/mL ampicillin and after overnight growth a single ampicillin-resistant colony is used to inoculate 10 mL of LB containing 100 μg/mL ampicillin. After growth for 10-16 hours with shaking (225 rpm) at 37°C, this culture is used to inoculate 1 L of LB containing 100 μg/mL ampicillin in a 1.5 L shaker flask. After growth at 37°C, 225 rpm, for 2 hours post-inoculation, the optical density at 600 nm is approximately 0.5 OD/mL. IPTG is added to 1mM and the culture allowed to grow for an additional 5-10 hours. Cells are harvested by centrifugation (5000 rpm, 10 minutes) and lysed by mechanical disruption. Recombinant human collagen is purified by ammonium sulfate fractionation and column chromatography. The yield is typically 15-25 mg/L of culture.

EXAMPLE 16

35 Hydroxyproline Incorporation Into Proteins In E. coli Under Proline Starvation Conditions

[0216] Seven plasmids, pGEX-4T.1 (Fig. 73), pTrc-TGF (Fig. 74), pMal-C2 (Fig. 1), pTrc-FN (Fig. 75), pTrc-FN-TGF (Fig. 76), pTrc-FN-Bmp (Fig. 77) and pGEX-HuColl^{Ec}, each separately containing genes encoding the following proteins: glutathione S-transferase (GST), the mature human TGF-β1 polypeptide (TGF-β1), mannose-binding protein (MBP), a 70 kDA fragment of human fibronectin (FN), a fusion of FN and TGF-β1 (FN-TGF-β1), a fusion of FN and human bone morphogenic protein 2A (FN-BMP-2A), and a fusion of GST and collagen (GST-Coll), were used individually to transform proline auxotrophic E. coli strain JM109 (F-). Transformation cultures were plated on LB agar containing 100 μg/ml ampicillin. After overnight incubation at 37°C, a single colony from a fresh transformation plate was used to inoculate 5 ml of LB media containing 400 mg ampicillin. After overnight growth at 37°C, this culture was centrifuged, the supernatant discarded, and the cell pellet washed twice with 5 ml of M9 medium (1X M9 salts, 0.5% glucose, 1 mM MgCl₂, 0.01% thiamine, 200 μg/ml glycine, 200 μg/ml alanine, 100 μg/ml of the other amino acids except proline, and 400 µg/ml ampicillin). The cells were finally resuspended in 5 ml of M9 medium. After incubation with shaking at 37°C for 30 minutes, trans-4-hydroxyproline was added to 40mM, NaCl to 0.5 M, and isopropyl-B-D-thiogalactopyyranoside to 1.5 mM. In certain cultures one of these additions was not made, as indicated in the labels for the lanes of the gels. After addition, incubation with shaking at 37°C was continued. After 4 hours, the cultures were centrifuged, the supernatants discarded, and the cell pellets resuspended in SDS-PAGE sample buffer (300 mM Tris (pH6.8)/0.5% SDS/10% glycerol/0.4M β-mercapthoethanol/0.2% bromophenol blue) to 15 OD600nm AU/ml, placed in boiling water bath for five minutes, and electrophoresed in denaturing polyacrylaminde gels. Proteins in the gels were visualized by staining with Coomassie Blue R250. The results of the gels are depicted in scans shown in Figs. 57-59. The scans relating to GST, TGF- β 1, MBP, FN, FN-TGF- β 1, and FN-BMP-2A (Figs. 57 and 58) show three lanes relating to each peptide, i.e., one lane indicating +NaCl/+Hyp wherein NaCl (hyperosmotic) and trans-4-hydroxyproline are present; one lane indicating -NaCl wherein trans-4-hydroxyproline is present but NaCl is not; and one lane indicating -Hyp which is +NaCl but absent trans-4-hydroxyproline. Asterisks on the scans mark protein bands which correspond

to the expressed target protein. The instances in which target protein was expressed all involve +NaCl in connection with +Hyp thus demonstrating +NaCl and +Hyp dependence.

[0217] The scan shown in Fig. 59 relating to GST-collagen shows four lanes relating to GST-Coll, i.e., one lane indicating +Hyp/+NaCl/-IPTG wherein *trans*-4-hydroxyproline and NaCl are present but IPTG (the protein expression inducer) is not and since there is no inducer, there is no target protein band; one lane indicating +NaCl/+IPTG/-Hyp wherein NaCl and IPTG are present but *trans*-4-hydroxyproline is not and, since *trans*-4-hydroxyproline is not present no target protein band is evident; one lane indicating +NaCl/+Pro/+IPTG wherein NaCl, proline and IPTG are present, but since the target protein is not stable when it contains proline, there is no target protein band; and one lane designated +IPTG/+NaCl/+Hyp wherein IPTG, NaCl and *trans*-4-hydroxyproline are present and since the protein is stabilized by the presence *of trans*-4-hydroxyproline an asterisk marked protein band is evident.

EXAMPLE 17

5

15

20

25

30

40

45

50:

55

Hydroxyproline incorporation into a collagen-like peptide in E. coli.

[0218] A plasmid (pGST-CM4, Figure 60) containing the gene for collagen mimetic 4 (CM4, Figure 61) (SEQ. ID. NO. 39) genetically linked to the 3' end of the gene for S. japonicum glutathione S-transferase was used to transform by electroporation proline auxotrophic E. coli strain JM109 (F-). Transformation cultures were plated on LB agar containing 100 μg/ml ampicillin. After overnight incubation at 37° C, a single colony from a fresh transformation plate was used to inoculate 5 ml of LB media containing 100 μg/ml ampicillin. After overnight growth at 37° C, 500 μl of this culture was centrifuged, the supernatent discarded, and the cell pellet washed once with 500 µl of M9 medium (1X M9 salts, 0.5 % glucose, 1 mM MgCl₂, 0.01 % thiamine, 200 µg/ml glycine, 200µg/ml alanine, 100 µg/ml of the other amino acids except proline, and 400 µg/ml ampicillin). The cells were finally suspended in 5 ml of M9 medium containing 10 μg/ml proline and 2 ml of this was used to inoculate 30 ml of M9 medium containing 10 μg/ml proline. After incubation with shaking at 37° C for 8 hours, the culture was centrifuged and the cell pellet washed once with M9 medium containing 5 μg/ml proline. The pellet was resuspended in 15 ml of M9 medium containing 5 μg/ml of proline and this culture was used to inoculate 1 L of M9 medium containing 5 μg/ml of proline. This culture was grown for 18 hours at 37° C to proline starvation. At this time, the culture was centrifuged, the cells washed once with M9 medium (with no proline), and the cells resuspended in 1 L of M9 medium containing 80 mM hydroxyproline, 0.5 M NaCl, and 1.5 mM isopropylβ-D-thiogalactopyranoside. Incubation was continued at 37° C with shaking for 22 hours. The cultures were centrifuged and the cell pellets stored at -20°C until processed further.

EXAMPLE 18

Proline incorporation into a collagen-like peptide in E. coli.

[0219] A plasmid (pGST-CM4, Figure 60) containing the gene for collagen mimetic 4 (CM4, Figure 61) (SEQ. ID. NO. 39) genetically linked to the 3' end of the gene for S. japonicum glutathione S-transferase was used to transform by electroporation proline auxotrophic E. coli strain JM109 (F-). Transformation cultures were plated on LB agar containing 100 μg/ml ampicillin. After overnight incubation at 37° C, a single colony from a fresh transformation plate was used to inoculate 5 ml of LB media containing 100 μg/ml ampicillin. After overnight growth at 37° C, 500 μl of this culture was centrifuged, the supernatent discarded, and the cell pellet washed once with 500 µl of M9 medium (1X M9 salts, 0.5 % glucose, 1 mM MgCl₂, 0.01 % thiamine, 200 μg/ml glycine, 200 μg/ml alanine, 100 μg/ml of the other amino acids except proline, and 400 µg/mL ampicillin). The cells were finally resuspended in 5 ml of M9 medium containing 10 μg/ml proline and 2 ml of this was used to inoculate 30 ml of M9 medium containing 10 μg/ml proline. This culture was incubated with shaking at 37° C for 8 hours. The culture was centrifuged and the cell pellet washed once with M9 medium containing 5 μg/ml proline. The pellet was resuspended in 15 ml of M9 medium containing 5 μg/ml of proline and this culture was used to inoculate 1 L of M9 medium containing 5 µg/ml of proline. This culture was grown for 18 hours at 37°C to proline starvation. At this time, the culture was centrifuged, the cells washed once with M9 medium (with no proline), and finally the cells were resuspended in 1 L of M9 medium containing 2.5 mM proline, 0.5 M NaCl, and 1.5 mM isopropyl-p-β-thiogalactopyranoside. Incubation was continued at 37° C with shaking for 22 hours. The cultures were then centrifuged and the cell pellets stored at -20°C until processed further.

EXAMPLE 19

Purification of hydroxyproline-containing collagen-like peptide from E. coli

[0220] The cell pellet from a 1 L fermentation culture prepared as described in Example 17 above, was resuspended

in 20 ml of Dulbecco's phosphate buffered saline (pH 7.1) (PBS) containing 1 mM EDTA, 100 μM PMSF, 0.5 μg/ml E64, and 0.7 μg/ml pepstatin (resuspension buffer). The cells were lysed by twice passing through a French press. Following lysis, the suspension was centrifuged for 30 minutes at 30,000 xg. The supernatent was discarded and the pellet washed once with 5 ml of resuspension buffer containing 1 M urea and 0.5% Triton X100 followed by one wash with 7 ml of resuspension buffer without urea or Triton X100. The pellet was finally resuspended in 5 ml of 6M guanidine hydrochloride in Dulbecco's phosphate buffered saline (pH7.1) containing 1 mM EDTA and 2 mM β-mercaptoethanol and sonicated on ice for 3 x 60 seconds (microtip, power = 3.5, Heat Systems XL-2020 model sonicator). The sonicated suspension was incubated at 4° C for 18 hours and then centrifuged at 14,000 rpm in a microcentrifuge. The supernatent (6 ml) was dialyzed (10,000 MWCO) against 4 x 4 L of distilled water at 4°C. The contents of the dialysis tubing were transferred to a 150 ml round bottom flask and lyophilized to dryness. The residue (~30 mg) was dissolved in 3 ml of 70% formic acid and 40 mg of cyanogen bromide was added. The flask was flushed once with nitrogen, evacuated, and allowed to stir for 18 hours at room temperature. The contents of the flask were taken to dryness in vacuo at room temperature, the residue resuspended in 5 ml of distilled water and evaporated to dryness again. This was repeated 2 times. The residue was finally dissolved in 2 ml of 0.2% trifluoroacetic acid (TFA). The trifluoroacetic acid-soluble material was applied in 100 μl aliquots to a Poros R2 column (4.6 mm x 100 mm) running at 5 ml/min. with a starting buffer of 98% 0.1% trifluoroacetic acid in water/2% 0.1 % TFA in acetonitrile. The hydroxyproline-containing protein was eluted with of gradient of 2% 0.1% TFA/acetonitrile to 40% 0.1% TFA/acetonitrile over 25 column volumes (Fig. 62A). The collagen-mimetic eluted between 18 and 23% 0.1% TFA/acetonitrile. Figure 62A is a chromatogram of the elution of hydroxyproline containing CM4 from a Poros RP2 column (available from Perseptive Biosystems, Framingham, MA). The arrow indicates the peak containing hydroxyproline containing CM4. Fractions were assayed by SDS-PAGE and collagen mimetic-containing fractions were pooled and lyophilized. Lyophilized material was stored at -20°

EXAMPLE 20

5

۵

15

20

25

30

35

45

50

Purification of proline-containing collagen-like peptide from E. coli

[0221] The cell pellet from a 500 ml fermentation culture prepared as described in Example 18 above, was resuspended in 20 ml of Dulbecco's phosphate buffered saline (pH 7.1) (PBS) containing 10 mM EDTA, 100 µM PMSF, 0.5 μg/ml E64, and 0.06 μg/ml aprotinin. Lysozyme (2 mg) was added and the suspension incubated at 4° C for 60 minutes. The suspension was sonicated for 5 x 60 seconds (microtip, power = 3.5, Heat Systems XL-2020 model sonicator). The sonicated suspension was centrifuged at 20,000 xg for 15 minutes. The supernatent was adjusted to 1% Triton X100 and incubated for 30 minutes at room temperature with 7 ml of glutathione sepharose 4B pre-equilibrated in PBS. The suspension was centrifuged at 500 rpm for 3 minutes. The supernatent decanted, and the resin washed 3 times with 8 ml of PBS. Bound proteins were eluted with 3 aliquots (2 ml each, 10 minutes gentle rocking at room temperature) of 10 mM glutathione in 50 mM Tris (pH 8.0). Eluants were combined and dialyzed (10,000 MWCO) against 3 x 4 L of distilled water at 4° C. The contents of the dialysis tubing were transferred to a 150 ml round bottom flask and lyophilized to dryness. The residue was dissolved in 3 ml of 70% formic acid and 4 mg of cyanogen bromide was added. The flask was flushed once with nitrogen evacuated, and allowed to stir for 18 hours at room temperature. The contents of the flask were taken to dryness in vacuo at room temperature, the residue resuspended in 5 ml of distilled water, and evaporated to dryness again. This was repeated 2 times. The residue was finally dissolved in 2 ml of 0.2% trifluoroacetic acid (TFA). The trifluoroacetic acid-soluble material was applied in 100 μl aliquots to a Poros R2 column (4.6 mm x 100 mm) running at 5 ml/min. with a starting buffer of 98% 0.1% trifluoroacetic acid in water/2% 0.1% TFA in acetonitrile. Bound protein was eluted with of gradient of 2% 0.1% TFA/acetonitrile to 40% 0.1% TFA/acetonitrile over 25 column volumes (Figure 62B). The collagen-mimetic eluted between 24 and 27% 0.1% TFA/acetonitrile. Figure 62B is a chromatogram of the elution of proline containing CM4 from a Poros RP2 column. The arrow indicates the peak containing proline containing CM4. Fractions were assayed by SDS-PAGE and collagen mimetic-containing fractions were pooled and lyophilized. Lyophilized material was stored at -20° C.

EXAMPLE 21

Amino acid analysis of hydroxyproline-containing collagen mimetic and proline-containing collagen mimetic.

[0222] Approximately 30 μg of purified hydroxyproline-containing collagen mimetic and proline-containing collagen mimetic prepared as described in Examples 19 and 20, respectively, were dissolved in 250 μl of 6N hydrochloric acid in glass ampules. The ampules were flushed two times with nitrogen, sealed under vacuum, and incubated at 110°C for 23 hours. Following hydrolysis, samples were removed from the ampules and taken to dryness in vacuo. The samples were dissolved in 15 μl of 0.1N hydrochloric acid and subjected to amino acid analysis on a Hewlett Packard

AminoQuant 1090 amino acid analyzer utilizing standard OPA and FMOC derivitization chemistry. Examples of the results of the amino acid analysis that illustrate the region of the chromatograms where the secondary amino acids (proline and hydroxyproline) elute are shown in Figures 63A through 63D. These Figures also show chromatograms of proline and hydroxyproline amino acid standards. More particularly, Figure 63A, depicts a chromatogram of a proline amino acid standard (250 pmol). *indicates a contaminating peak; Figure 63B depicts a chromatogram of a hydroxyproline amino acid standard (250 pool). *indicates a contaminating peak. Figure 63C depicts an amino analysis chromatogram of the hydrolysis of proline-containing CM4. Only the region of the chromatogram where proline and hydroxyproline elute is shown. *indicates a contaminating peak. Figure 63D depicts an amino acid analysis chromatogram of the hydrolysis of hydroxyproline-containing CM4. Only the region of the chromatogram where proline and hydroxyproline elute is shown. *indicates a contaminating peak.

EXAMPLE 22

5

10

ί.

15

25

30

ਂ 35

40

45

55

Determination of proline starvation conditions for E. coli (strain JM109 (F-))

[0223] A plasmid (pGST-CM4, Figure 60) containing the gene for collagen mimetic 4 (CM4, Figure 61) genetically linked to the 3' end of the gene for S. japonicum glutathione S-transferase was used to transform by electroporation proline auxotrophic E. coli strain JM109 (F-). Transformation cultures were plated on LB agar containing 100 µg/ml ampicillin. After overnight incubation at 37 °C, a single colony from a fresh transformation plate was used to inoculate 2 ml of M9 media (1X M9 salts, 0.5 % glucose, 1 mM MgCl₂, 0.01 % thiamine, 200 μg/ml glycine, 200 μg/ml alanine, 100 μg/ml of the other amino acids except proline, and 200 μg/ml carbenicillin) and containing 20 μg/ml proline. After growth at 37° C with shaking for 8 hours, 1.5 ml was used to inoculate 27 ml of M9 media containing 45 µg/ml proline. After incubation at 37° C with shaking for 7 hours, the culture was centrifuged, the cell pellet washed with 7 ml of M9 media with no proline, and finally resuspended in 17 ml of M9 media with no proline. This culture was used to inoculate four 35 mt cultures of M9 media containing 4 µg/mt proline at an OD600 of 0.028. Cultures were incubated with shaking at 37° C and the OD600 monitored. After 13.5 hours growth, the OD600 had plateaued. At this time, one culture was supplemented with proline at 15 μg/ml, one with hydroxyproline at 15 μg/ml, one with all of the amino acids at 15 μg/ ml except proline and hydroxyproline, and one culture with nothing. Incubation was continued and the OD600 monitored for a total of 24 hours. Figure 64 is a graph of OD600 vs. time for cultures of JM109 (F-) grown to plateau and then supplemented with various amino acids. The point at which the cultures were supplemented is indicated with an arrow. Proline starvation is evident since only the culture supplemented with proline continued to grow past plateau.

EXAMPLE 23

Hydroxyproline Incorporation Into Type I (α1) Collagen in E. coli

[0224] A plasmid (pHuCol(α 1)^{Ec}, Figure 65) containing the gene for Type I (α 1) collagen with optimized E. coli codon usage (Figure 39A-39E) (SEQ. ID. NO. 19) under control of the tac promoter and containing the gene for chloramphenicol resistance was used to transform by electroporation proline auxotrophic E. coli strain JM109 (F-). Transformation cultures were plated on LB agar containing 20 μg/ml chloramphenicol. After overnight incubation at 37 °C, a single colony from a fresh transformation plate was used to inoculate 100 ml of LB media containing 20 μg/ml chloramphenicol. This culture was grown to an OD600nm of 0.5 and 100 μl aliquots transferred to 1.5 ml tubes. The tubes were stored at -80 ° C. For expression, a tube was thawed on ice and used to inoculate 25 ml of LB media containing 20 μg/ml chloramphenicol. After overnight growth at 37° C, a four ml aliquot was withdrawn, centrifuged, the cell pellet washed once with 1 ml of 2x YT media containing 20 μg/ml chloramphenicol, and the washed cells used to inoculate 1 L of 2x YT medium containing 20 μg/ml chloramphenicol. This culture was grown at 37° C to an OD600nm of 0.8. The culture was centrifuged and the cell pellet washed once with 100 ml of M9 medium (1X M9 salts, 0.5 % glucose, 1 mM MgCl₂, 0.01 % thiamine, 200 μg/ml glycine, 200 μg/ml alanine, 100 μg/ml of the other amino acids except proline, and 20 µg/ml chloramphenicol). The cells were resuspended in 910 ml of M9 medium (1X M9 salts, 0.5 % glucose, 1 mM MgCl₂, 0.01 % thiamine, 200 μg/ml glycine, 200 μg/ml alanine, 100 μg/ml of the other amino acids except proline, and 20 µg/ml chloramphenicol) and allowed to grow at 37° C for 30 minutes. NaCl (80 ml of 5 M), hydroxyproline (7.5 ml of 2M), and IPTG (500 µl of 1 M) were added and growth continued for 3 hours. Cells were harvested by centrifugation and stored at -20° C.

EXAMPLE 24

5

10

15

20

30

40

50

Hydroxyproline Incorporation Into Type I (α2) in E. coli

[0225] A plasmid (pHuCol(α 2)^{Ec}, Figure 66) containing the gene for Type I (α 2) collagen with optimized E. coli codon usage (Figure 50A-50E) (SEQ. ID. NO. 31) under control of the tac promoter and containing the gene for chloramphenicol resistance was used to transform by electroporation proline auxotrophic E. coli strain JM109 (F-). Transformation cultures were plated on LB agar containing 20 µg/ml chloramphenicol. After overnight incubation at 37° C, a single colony from a fresh transformation plate was used to inoculate 100 ml of LB media containing 20 µg/ml chloramphenicol. This culture was grown to an OD600nm of 0.5 and 100 µl aliquots transferred to 1.5 ml tubes. The tubes were stored at -80 ° C. For expression, a tube was thawed on ice and used to inoculate 25 ml of LB media containing 20 μg/ml chloramphenicol. After overnight growth at 37° C, a four ml aliquot was withdrawn, centrifuged, the cell pellet washed once with 1 ml of 2x YT media containing 20 µg/ml chloramphenicol, and the washed cells used to inoculate 1 L of 2x YT medium containing 20 µg/ml chloramphenicol. This culture was grown at 37° C to an OD600nm of 0.8. The culture was centrifuged and the cell pellet washed once with 100 ml of M9 medium (1X M9 salts, 0.5 % glucose, 1 mM MgCl₂, 0.01 % thiamine, 200 μg/ml glycine, 200 μg/ml alanine, 100 μg/ml of the other amino acids except proline, and 20 µg/ml chloramphenicol). The cells were resuspended in 910 ml of M9 medium (1X M9 salts, 0.5 % glucose, 1 mM MgCl₂, 0.01 % thiamine, 200 μg/ml glycine, 200 μg/ml alanine, 100 μg/ml of the other amino acids except proline, and 20 µg/ml chloramphenicol) and allowed to grow at 37° C for 30 minutes. NaCl (80 ml of 5 M), hydroxyproline (7.5 ml of 2M), and IPTG (500 µl of 1 M) were added and growth continued for 3 hours. Cells were harvested by centrifugation and stored at -20° C.

EXAMPLE 25

25 Hydroxyproline Incorporation Into a C-terminal Fragment of Type I (α1) Collagen in E. coli

[0226] A plasmid (pD4-α1, Figure 67) encoding the gene for the carboxy terminal 219 amino acids of human Type I (α1) collagen with optimized E. coli codon usage fused to the 3'-end of the gene for glutathione S-transferase and under control of the tac promoter and containing the gene for ampicillin resistance was used to transform by electroporation proline auxotrophic E. coli strain JM109 (F-). Transformation cultures were plated on LB agar containing 100 μg/ml ampicillin. After overnight incubation at 37° C, a single colony from a fresh transformation plate was used to inoculate 100 ml of LB media containing 100 μg/ml ampicillin. This culture was grown to an OD600nm of 0.5 and 100 μl aliquots transferred to 1.5 ml tubes. The tubes were stored at -80° C. For expression, a tube was thawed on ice and used to inoculate 25 ml of LB media containing 400 µg/ml ampicillin. After overnight growth at 37° C, a four ml aliquot was withdrawn, centrifuged, the cell pellet washed once with 1 ml of 2x YT media containing 400 µg/ml ampicillin, and the washed cells used to inoculate 1 L of 2x YT medium containing 400 μg/ml ampicillin. This culture was grown at 37° C to an OD600nm of 0.8. The culture was centrifuged and the cell pellet washed once with 100 ml of M9 medium (1X M9 salts, 0.5 % glucose, 1 mM MgCl₂, 0.01 % thiamine, 200 µg/ml glycine, 200 µg/ml alanine, 100 µg/ml of the other amino acids except proline, and 400 µg/ml ampicillin). The cells were resuspended in 910 ml of M9 medium (1X M9 salts, 0.5 % glucose, 1 mM MgCl₂, 0.01 % thiamine, 200 μg/ml glycine, 200 μg/ml alanine, 100 μg/ml of the other amino acids except proline, and 400 µg/ml ampicillin) and allowed to grow at 37° C for 30 minutes. NaCl (80 ml of 5 M), hydroxyproline (7.5 ml of 2M), and IPTG (500 pl of 1 M) were added and growth continued for 3 hours. Cells were harvested by centrifugation and stored at -20° C.

45 EXAMPLE 26

Hydroxyproline Incorporation Into a C-terminal Fragment of Type I (α2) Collagen in E. coli

[0227] A plasmid (pD4- α 2, Figure 68) encoding the gene for the carboxy terminal 219 amino acids of human Type I (α 2) collagen with optimized *E. coli* codon usage as constructed in accordance with Example 14A fused to the 3'-end of the gene for glutathione *S*-transferase and under control of the *tac* promoter and containing the gene for ampicillin resistance was used to transform by electroporation proline auxotrophic *E. coli* strain JM109 (F-). Transformation cultures were plated on LB agar containing 100 μ g/ml ampicillin. After overnight incubation at 37° C, a single colony from a fresh transformation plate was used to inoculate 100 ml of LB media containing 100 μ g/ml ampicillin. This culture was grown to an OD600nm of 0.5 and 100 μ l aliquots transferred to 1.5 ml tubes. The tubes were stored at -80° C. For expression, a tube was thawed on ice and used to inoculate 25 ml of LB media containing 400 μ g/ml ampicillin. After overnight growth at 37° C, a four ml aliquot was withdrawn, centrifuged, the cell pellet washed once with 1 ml of 2x YT media containing 400 μ g/ml ampicillin, and the washed cells used to inoculate 1 L of 2x YT medium containing

400 μ g/ml ampicillin. This culture was grown at 37° C to an OD600nm of 0.8. The culture was centrifuged and the cell pellet washed once with 100 ml of M9 medium (1X M9 salts, 0.5 % glucose, 1 mM MgCl₂, 0.01 % thiamine, 200 μ g/ml glycine, 200 μ g/ml alanine, 100 μ g/ml of the other amino acids except proline, and 400 μ g/ml ampicillin). The cells were resuspended in 910 ml of M9 medium (1X M9 salts, 0.5 % glucose, 1 mM MgCl₂, 0.01 % thiamine, 200 μ g/ml glycine, 200 μ g/ml alanine, 100 μ g/ml of the other amino acids except proline, and 400 μ g/ml ampicillin) and allowed to grow at 37° C for 30 minutes. NaCl (80 ml of 5 M), hydroxyproline (7.5 ml of 2M), and IPTG (500 μ l of 1 M) were added and growth continued for 3 hours. Cells were harvested by centrifugation and stored at -20° C.

EXAMPLE 27

10

15

20

25

30

45

50

55

Purification of Hydroxyproline-containing C-terminal Fragment of Type I (α 1) Collagen

[0228] Cell paste harvested from a 1 L culture grown as in Example 25 was resuspended in 30 ml of lysis buffer (2M urea, 137mM NaCl, 2.7mM KCl, 4.3mM Na $_2$ HPO $_4$, 1.4mM KH $_2$ PO $_4$, 10mM EDTA, 10mM βME, 0.1% Triton X-100, pH 7.4) at 4°C. Lysozyme (chicken egg white) was added to 100 μg/ml and the solution incubated at 4 °C for 30 minutes. The solution was passed twice through a cell disruption press (SLM Instruments, Rochester, NY) and then centrifuged at 30,000 x g for 30 minutes. The pellet was resuspended in 30 ml of 50 mM Tris-HCl, pH 7.6, centrifuged at 30,000 x g for 30 minutes, and the pellet solubilized in 25 ml of solubilization buffer (8M urea, 137mM NaCl, 2.7mM KCl, 4.3mM Na $_2$ HPO $_4$, 1.4mM KH $_2$ PO $_4$, 5mM EDTA, 5mM βME). The solution was centrifuged at 30,000xg for 30 minutes and supernatent dialyzed against two changes of 4 L of distilled water at 4°C. Following dialysis, the entire mixture was lyophilized. The lyophilized solid was dissolved in 0.1M HCl in a flask with stirring. After addition of a 5-fold excess of crystalline BrCN, the flask was evacuated and filled with nitrogen. Cleavage was allowed to proceed for 24 hrs, at which time the solvent was removed in vacuo. The residue was dissolved in 0.1% trifluoroacetic acid (TFA) and purified by reverse-phase HPLC using a Vydac C4 RP-HPLC column (10x250mm, 5μ, 300 Å) on a BioCad Sprint system (Perceptive Biosystems, Framingham, MA). Hydroxyproline-containing D4 protein was eluted with a gradient of 15-40% acetonitrile/0.1% TFA over a 45 minute period. Protein D4-α1 eluted at 26% acetonitrile/0.1% TFA.

EXAMPLE 28

Purification of Hydroxyproline-containing C-terminal Fragment of Type I (α2) Collagen

[0229] Cell paste harvested from a 1 L culture grown as in Example 26 was resuspended in 30 ml of lysis buffer (2M urea, 137mM NaCl, 2.7mM KCl, 4.3mM Na₂HPO₄, 1.4mM KH₂PO₄, 10mM EDTA, 10mM βME, 0.1% Triton X-100, pH 7.4) at 4°C. Lysozyme (chicken egg white) was added to 100 μg/ml and the solution incubated at 4°C for 30 minutes. The solution was passed twice through a cell disruption press (SLM Instruments, Rochester, NY) and then centrifuged at 30,000 x g for 30 minutes. The pellet was resuspended in 30 ml of 50 mM Tris-HCl, pH 7.6, centrifuged at 30,000 x g for 30 minutes, and the pellet solubilized in 25 ml of solubilization buffer (8M urea, 137mM NaCl, 2.7mM KCl, 4.3mM Na₂HPO₄, 1.4mM KH₂PO₄, 5mM EDTA, 5mM βME). The solution was centrifuged at 30,000xg for 30 minutes and supernatent dialyzed against two changes of 4 L of distilled water at 4°C. Following dialysis, the entire mixture was lyophilized. The lyophilized solid was dissolved in 0.1 M HCl in a flask with stirring. After addition of a 5-fold excess of crystalline BrCN, the flask was evacuated and filled with nitrogen. Cleavage was allowed to proceed for 24 hrs, at which time the solvent was removed in vacuo. The residue was dissolved in 0.1% trifluoroacetic acid (TFA) and purified by reverse-phase HPLC using a Vydac C4 RP-HPLC column (10x250mm, 5μ, 300 Å) on a BioCad Sprint system (Perceptive Biosystems, Framingham, MA). Hydroxyproline-containing D4 protein was eluted with a gradient of 15-40% acetonitrile/0.1 % TFA over a 45 minute period. Protein D4- α 2 eluted at 25% acetonitrile/0.1 % TFA.

EXAMPLE 29

Amino Acid Composition Analysis of Hydroxyproline-containing C-terminal Fragment of Type I (α1) Collagen

[0230] Protein D4-α1 (10μg) purified as in Example 27 was taken to dryness in vacuo in a 1.5 ml microcentrifuge tube. A sample was subjected to amino acid analysis at the W.M. Keck Foundation Biotechnology Resource Laboratory (New Haven, CT) on an Applied Biosystems sequencer equipped with an on-line HPLC system. The experimentally determined sequence of the first 13 amino acids (SEQ. ID. NO. 41) and the sequence predicted from the DNA sequence (SEQ. ID. NO. 42) are shown in Figure 69. A sample of protein D4-al was subjected to mass spectral analysis on a VG Biotech BIO-Q quadrople analyzer at M-Scan, Inc. (West Chester, PA). The mass spectrum and the predicted molecular weight of protein D4-α1 if it contained 100% hydroxyproline in lieu of proline are given in Figure 70. The predicted molecular weight of protein D4-α1 containing 100% hydroxyproline in lieu of proline is 20807.8 Da. The

experimentally determined molecular weight was 20807.5 Da.

EXAMPLE 30

5

10

15

20

30

40

45

Construction of Carboxy Terminal 219 Amino Acids of Human Collagen Type I (α1) Fragment Gene with Optimized E. Coli Codon Usage.

[0231] The nucleotide sequence of the 657 nucleotide gene for the carboxy terminal 219 amino acids of human Type I (α 1) collagen with optimized *E. Coli* codon usage is shown in Figure 71. For synthesis of this gene, unique restriction sites were identified or created approximately every 150 base pairs. Oligos of approximately 80 nucleotides were synthesized on a Beckman Oligo 1000 DNA synthesizer, cleaved and deprotected with aqueous NH₄OH, and purified by electrophoresis in 7M urea/12% polyacrylamide gels. Each set of oligos was designed to have an EcoR I restriction enzyme site at the 5' end, a unique restriction site near the 3' end, followed by the TAAT stop sequence and a Hind III restriction enzyme site at the very 3' end. The first four oligos, comprising the first 84 amino acids of the carboxy terminal 219 amino acids of human Type I (α 1) collagen with optimized *E. coli* codon usage, are given in Figure 81 (SEQ. ID. NOS. 47-50).

[0232] Oligos N4-1 (SEQ. ID. NO. 47) and N4-2 (SEQ. ID. NO. 48) (1 μg each) were annealed in 20 μL of T7 DNA polymerase buffer (40mM Tris-HCI (pH 8.0), 5mM MgCl₂, 5mM dithiothreitol, 50mM NaCl, 0.05 mg/mL bovine serum albumin) by heating at 90°C for 5 minutes followed by slow cooling to room temperature. After brief centrifugation at 14,000 rpm, 10 units of T7 DNA polymase and 2 µL of a solution of all four dNTPs (dATP, dGTP, dCTP, dTTP, 2.5mM each) were added to the annealed oligos. Extension reactions were incubated at 37°C for 30 minutes and then heated at 70°C for 10 minutes. After cooling to room temperature, Hind III buffer (5 µL of 10 x concentration), 20 µL of H₂O, and 10 units of Hind III restriction enzyme were added and the tubes incubated at 37°C for 10 hours. Hind III buffer (2 μL of 10x concentration), 13.5 μL of 0.5M Tris HCl (pH 7.5), 1.8 μL of 1% Triton X100, 5.6 μL of H20, and 20 U of EcoR I were added to each tube and incubation continued for 2 hours at 37°C. Digests were extracted once with an equal volume of phenol, once with phenol/chloroform/isoamyl alcohol, and once with chloroform/isoamyl alcohol. After ethanol precipitation, the pellet was resuspended in 10 µL of TE buffer (10 mM Tris HCI (pH 8.0), 1 mM EDTA). Resuspended pellet 4 μL of was ligated overnight at 16°C with agarose gel-purified EcoRI/Hind III digested pBSKS+ vector (1 μg) using T4 DNA ligase (100 units). One half of the transformation mixture was transformed by heat shock into DH5α cells and 100 μL of the 1.0 mL transformation mixture was plated on Luria Broth (LB) agar plates containing 70 μg/mL ampicillin. Plates were incubated overnight at 37°C. Ampicillin resistant colonies (6-12) were picked and grown overnight in LB media containing 70µg/mL ampicillin. Plasmid DNA was isolated from each culture by Wizard Minipreps (Promega Corporation, Madison WI) and screened for the presence of the approximately 120 base pair insert by digestion with EcoRI and Hind III and running the digestion products on agarose electrophoresis gels. Clones with inserts were confirmed by standard dideoxy termination DNA sequencing. The correct clone was named pBSN4-1.

[0233] Oligos N4-3 (SEQ. ID. NO. 49) and N4-4 (SEQ. ID. NO. 50) (Figure 81) were synthesized, purified, annealed, extended, and cloned into pBSKS+ following exactly the same procedure given above for oligos N4-1 and N4-2. The resulting plasmid was named pBSN4-2A. To clone together the sections of the collagen gene from pBSN4-1 and pBSN4-2A, plasmid pBSN4-1 (1μg) was digested for 2 hours at 37°C with Apa L1 and Hind III. The digested vector was purified by agarose gel electrophoresis. Plasmid pBSN4-2A (3 μg) was digested for 2 hours at 37°C with Apa L1 and Hind III and the insert purified by agarose gel electrophoresis. Apa L1/Hind III-digested pBSN4-1 was ligated with this insert overnight at 16°C with T4 DNA ligase. One half of the ligation mixture was transformed into DH5α cells and 1/10 of the transformation mixture was plated on LB agar plates containing 70 μg/mL ampicillin. After overnight incubation at 37°C, ampicillin-resistant clones were picked and screened for the presence of insert DNA as described above. Clones were confirmed by dideoxy termination sequencing. The correct clone was named pBSN4-2.

[0234] In a similar manner, the remainder of the gene for the carboxy terminal 219 amino acids of human Type I (α 1) collagen with optimized *E. coli* codon usage was constructed such that the final DNA sequence is that given in Figure 71 (SEQ. ID. NO. 43).

[0235] It will be understood that various modifications may be made to the embodiments disclosed herein. For example, it is contemplated that any protein produced by prokaryotes and eukaryotes can be made to incorporate one or more amino acid analogs in accordance with the present disclosure. Therefore, the above description should not be construed as limiting, but merely as exemplifications of preferred embodiments. Those skilled in art will envision other modifications within the scope and spirit of the claims appended hereto.

Annex to the description

[0236]

5

SEQUENCE LISTING

10	(1) GENERAL INFORMATION:
15	(i) APPLICANT: GRUSKIN, ELLIOT A. BUECHTER, DOUGLAS
	BROKAW, JANE
	ZHANG, GUANGHUI
20	PAOLELLA, DAVID
25	(ii) TITLE OF INVENTION: AMINO ACID MODIFIED POLYPEPTIDES
25	(iii) NUMBER OF SEQUENCES: 50
, 30	(iv) CORRESPONDENCE ADDRESS:
30	(A) ADDRESSEE: DILWORTH & BARRESE
	(B) STREET: 333 EARLE OVINGTON BOULEVARD
£ .	(C) CITY: UNIONDALE
35	(D) STATE: NY
	(E) COUNTRY: U.S.A.
	(F) ZIP: 11553
40	
	(v) COMPUTER READABLE FORM:
	(A) MEDIUM TYPE: Floppy disk
45	(B) COMPUTER: IBM PC compatible
	(C) OPERATING SYSTEM: PC-DOS/MS-DOS
	(D) SOFTWARE: PatentIn Release #1.0, Version #1.30
50	
	(vi) CURRENT APPLICATION DATA:
	(A) APPLICATION NUMBER:
55	(B) FILING DATE:
	(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

	(A) NAME: STEEN, JEFFREY S	
5		
	(ix) TELECOMMUNICATION INFORMATION:	
	(A) TELEPHONE: (516) 228-8484	
0	(B) TELEFAX: (516) 228-8516	
5	(2) INFORMATION FOR SEQ ID NO:1:	
	(i) SEQUENCE CHARACTERISTICS:	
20	(A) LENGTH: 3170 base pairs	
	(B) TYPE: nucleic acid	•
	(C) STRANDEDNESS: single (D) TOPOLOGY: linear	
25	(b) lorologi: linear	
	(ii) MOLECULE TYPE: cDNA	
	(22, 33223333333333333333333333333333333	
30 ·	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:	
,-		
	CAGCTGTCTT ATGGCTATGA TGAGAAATCA ACCGGAGGAA TTTCCGTGCC TGGCCCCATG	60
35		
	GGTCCCTCTG GTCCTCGTGG TCTCCCTGGC CCCCCTGGTG CACCTGGTCC CCAAGGCTTC	120
40	CAAGGTCCCC CTGGTGAGCC TGGCGAGCCT GGAGCTTCAG GTCCCATGGG TCCCCGAGGT	180
	CCCCCAGGTC CCCCTGGAAA GAATGGAGAT GATGGGGAAG CTGGAAAACC TGGTCGTCCT	240
	CCCCAGGIC CCCCIOGNA GARIGONGAI GAIGGGGAAG CIGGILLEGO IGGICCICGI	
4 5	GGTGAGCGTG GGCCTCCTGG GCCTCAGGGT GCTCGAGGAT TGCCCGGAAC AGCTGGCCTC	300
:		
	CCTGGAATGA AGGGACACAG AGGTTTCAGT GGTTTGGATG GTGCCAAGGG AGATGCTGGT	360
50		
4	CCTGCTGGTC CTAAGGGTGA GCCTGGCAGC CCTGGTGAAA ATGGAGCTCC TGGTCAGATG	420
-		

480	TGCTGGTGCT	CCCCTGGCCC	CGCCCTGGAG	TGAGAGAGGT	GCCTGCCTGG	GGCCCCCGTG
540	CCCCGCTGGT	GTCCCACCGG	eeecccccie	TGGTGCTGCC	ATGGTGCTAC	CGTGGAAATG
600	AGGGCCCCGA	CTGGTCCCCA	AAGGGTGAAG	TGTTGGTGCT	TCCCTGGTGC	CCTCCTGGCT
660	TGCTGGTGCT	CCCCTGGCCC	GAGCCTGGCC	TGTGCGTGGT	GTCCCCAGGG	GGCTCTGAAG
720	TGCCAATGGT	GTGCTAAAGG	GGACAGCCTG	TGGTGCTGAT	CTGGAAACCC	GCTGGCCCTG
780	TGGACCCCAG	GAGGCCCCTC	CCTGGTGCCC	TCCTGGCTTC	TTGCTGGTGC	GCTCCTGGTA
840	TCCTGGCAGC	AACCTGGTGC	AACAGCGGTG	TCCCAAGGGT	GCCCTCCTGG	GGCCCCGGCG
900	ACCCCCTGGC	GTGTTCAAGG	GGCCCTGTTG	GGGAGAGCCT	CTGGTGCTAA	AAAGGAGACA
960	TGGCCTGCCC	CCGGACCCAC	CGAGGTGAAC	GCGAGGAGCT	AGGAAGGAAA	CCTGCTGGAG
1020	AGATGGTGTT	TCCCTGGCGC	AGCCGTGGTT	TGGACCTGGT	GCGAGCGTGG	GGACCCCCTG
1080	CCCCAAAGGA	GCCCCGCTGG	GGTTCTCCTG	TGGTGAACGT	AGGGTCCCGC	GCTGGTCCCA
1140	GGGTCTGACT	CTGGTGCCAA	GCTGGTCTGC	TCCCGGTGAA	AAGCTGGTCG	TCTCCTGGTG
1200	CGCCGGTCAA	CCCCTGGTCC	AAAACTGGCC	TCCTGATGGC	GCAGCCCTGG	GGAAGCCCTG
1260	TGTGATGGGA	GTCAGGCTGG	GGTGCCCGTG	AGGCCCACCT	CCGGACCCCC	GATGGTCGCC
1320	AGGTGTTCCC	CTGGAGAGCG	CCCGGCAAGG	TGCTGGAGAG	CTAAAGGTGC	TTCCCTGGAC
1380	TCAGGGACCC	AGGCTGGAGC	AAAGATGGAG	TCCTGCTGGC	GCGCTGTCGG	GGACCCCCTG
1440	CTCCCCCGA	GCCCTGCTGG	GGTGAACAAG	TGGCGAGAGA	CTGGTCCCGC	ССТСССССТС

.

	TTCCAGGGTC	TCCCTGGTCC	TGCTGGTCCT	CCAGGTGAAG	CAGGCAAACC	TGGTGAACAG	1500
5	GGTGTTCCTG	GAGACCTTGG	CGCCCCTGGC	CCCTCTGGĄG	CAAGAGGCGA	GAGAGGTTTC	1560
10	CCTGGCGAGC	GTGGTGTGCA	AGGTCCCCCT	GGTCCTGCTG	GACCCCGAGG	GGCCAACGGT	1620
10	GCTCCCGGCA	ACGATGGTGC	TAAGGGTGAT	GCTGGTGCCC	CTGGAGCTCC	CGGTAGCCAG	1680
15	GGCGCCCCTG	GCCTTCAGGG	AATGCCTGGT	GAACGTGGTG	CAGCTGGTCT	TCCAGGGCCT	1740
	AAGGGTGACA	GAGGTGATGC	TGGTCCCAAA	GGTGCTGATG	GCTCTCCTGG	CAAAGATGGC	1800
20	GTCCGTGGTC	TGACCGGCCC	CATTGGTCCT	CCTGGCCCTG	CTGGTGCCCC	TGGTGACAAG	1860
25	GGTGAAAGTG	GTCCCAGCGG	CCCTGCTGGT	CCCACTGGAG	CTCGTGGTGC	CCCCGGAGAC	1920
	CGTGGTGAGC	CTGGTCCCCC	CGGCCCTGCT	GGCTTTGCTG	GCCCCCTGG	TGCTGACGGC	1980
30	CAACCTGGTG	CTAAAGGCGA	ACCTGGTGAT	GCTGGTGCCA	AAGGCGATGC	TGGTCCCCT	2040
	GGGCCTGCCG	GACCCGCTGG	ACCCCCTGGC	CCCATTGGTA	ATGTTGGTGC	TCCTGGAGCC	2100
35	AAAGGTGCTC	GGGCAGCGCT	GGTCCCCCTG	GTGCTACTGG	TTTCCCTGGT	GCTGCTGGCC	2160
; 40	GAGTCGGTCC	TCCTGGCCCC	TCTGGAAATG	CTGGACCCCC	TGGCCCTCCT	GGTCCTGCTG	2220
	GCAAAGAAGG	CGGCAAAGGT	CCCCGTGGTG	AGACTGGCCC	TGCTGGACGT	. CCTGGTGAAG	2280
45 [:]	TTGGTCCCCC	TGGTCCCCCT	GGCCCTGCTG	GCGAGAAAGG	ATCCCCTGGT	GCTGATGGTC	2340
	CTGCTGGTGC	TCCTGGTACT	CCCGGGCCTC	AAGGTATTGC	TGGACAGCGT	GGTGTGGTCG	2400
50	GCCTGCCTGG	TCAGAGAGGA	GAGAGAGGCT	TCCCTGGTCT	TCCTGGCCCC	TCTGGTGAAC	2460

CTGGCAAACA	AGGTCCCTCT	GGAGCAAGTG	GTGAACGTGG	TCCCCCGGT	CCCATGGGCC	2520
CCCCTGGATT	GGCTGGACCC	CCTGGTGAAT	CTGGACGTĢA	GGGGGCTCCT	GCTGCCGAAG	2580
GTTCCCCTGG	ACGAGACGGT	TCTCCTGGCG	CCAAGGGTGA	CCGTGGTGAG	ACCGGCCCCG	2640
CTGGACCCCC	TGGTGCTCCT	GGTGCTCCTG	GTGCCCCTGG	CCCCGTTGGC	CCTGCTGGCA	2700
AGAGTGGTGA	TCGTGGTGAG	ACTGGTCCTG	CTGGTCCCGC	CGGTCCCGTC	GGCCCCGCTG	2760
GCGCCCGTGG	CCCCGCCGGA	CCCCAAGGCC	CCCGTGGTGA	CAAGGGTGAG	ACAGGCGAAC	2820
AGGGCGACAG	AGGCATAAAG	GGTCACCGTG	GCTTCTCTGG	CCTCCAGGGT	CCCCTGGCC	2880
CTCCTGGCTC	TCCTGGTGAA	CAAGGTCCCT	CTGGAGCCTC	TGGTCCTGCT	GGTCCCCGAG	2940
GTCCCCCTGG	CTCTGCTGGT	GCTCCTGGCA	AAGATGGACT	CAACGGTCTC	CCTGGCCCCA	3000
TTGGGCCCCC	TGGTCCTCGC	GGTCGCACTG	GTGATGCTGG	TCCTGTTGGT	cccccaacc	3060
CTCCTGGACC	TCCTGGTCCC	CCTGGTCCTC	CCAGCGCTGG	TTTCGACTTC	AGCTTCCTCC	3120
CCCAGCCACC	TCAAGAGAAG	GCTCACGATG	GTGGCCGCTA	CTACCGGGCT		3170

(2) INFORMATION FOR SEQ ID NO:2:

40 (i) SEQUENCE CHARACTERISTICS:

5

10

15

20

25

30

35

45

50

55

(A) LENGTH: 240 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:	
.£ .5	CAGCTGTCTT ATGGCTATGA TGAGAAATCA ACCGGAGGAA TTTCCGTGCC TGGCCCCATG	60
ੱ। 10	GGTCCCTCTG GTCCTCGTGG TCTCCCTGGC CCCCCTGGTG CACCTGGTCC CCAAGGCTTC	120
	CAAGGTCCCC CTGGTGAGCC TGGCGAGCCT GGAGCTTCAG GTCCCATGGG TCCCCGAGGT	180
 15	CCCCCAGGTC CCCCTGGAAA GAATGGAGAT GATGGGGAAG CTGGAAAACC TGGTCGTCCT	240
20	(2) INFORMATION FOR SEQ ID NO:3:	
	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 100 base pairs	
25	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
.: 30		
30	(ii) MOLECULE TYPE: cDNA	
హ 35	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:	
	GGATCCATGG GGCTCGCTGG CCCACCGGGC GAACCGGGTC CGCCAGGCCC GAAAGGTCCG	60
40	CGTGGCGATA GCGGGCTCCC GGGCGATTCC TAATGGATCC	100

;

,	(2)	INFO	RMATI	ON FOR	SEQ ID	NO:4:									
5		(4)	SROIT	ENCE CH	ARACTE	PISTIC	c.								
		(4)		LENGTH				•							
				TYPE:			CIUB								
10				STRAND			10								
				TOPOLO			16								
		•	(D)	TOPOLO	G1. un	KIIOWII									
15		(ii)	MOLE	CULE TY	PE: pe	ptide									
•		(xi)	SEQU	ENCE DE	SCRIPT	'ION: S	EQ ID NO	:4:							
20		0 1	7	cl.	. D			a 1	D	D	a 1	n			
,			Leu .	Ala Giy		ro Gly	Glu Pro		Pro	PIO	GIÀ	Pro	•	GIÀ	•
		1			5			10					15		
25		Pro	Arg	Gly Asp	Ser										
				20											
30	(2)	INFO	RMATI	ON FOR	SEQ ID	NO:5:									
		(i)	SEQU	ENCE CH	ARACTE	RISTIC	S:								
35			(A)	LENGTH	: 330	base p	airs								
			(B)	TYPE:	nuclei	c acid	l								
			(C)	STRAND	EDNESS	: sing	le								
40 ⁻			(D)	TOPOLO	GY: li	near									
		(ii)	MOLE	CULE TY	PE: cD	NA									
45		(xi)	SEQU	ENCE DE	SCRIPT	'ION: S	EQ ID NO	:5:							
50	CAG	cggc	CA GG	AAGAAGA	A TAAG	AACTGC	ceecec	ACT (CGCT	CTATG	T GG	ACTI	CAGO	2	60
	GAT	GTGGG	CT GG	AATGACT	G GATT	'GTGGCC	CCACCAG	GCT A	ACCAC	GCCI	T CI	ACTO	CCAT	r	120

42

<

	GGGGACTGCC CCTTTCCACT GGCTGACCAC CTCAACTCAA	180
5	ACCCTGGTCA ATTCTGTCAA TTCCAGTATC CCCAAAGCCT GTTGTGTGCC CACTGAACTG	240
10	AGTGCCATCT CCATGCTGTA CCTGGATGAG TATGATAAGG TGGTACTGAA AAATTATCAG	300
	GAGATGGTAG TAGAGGGATG TGGGTGCCGC	330
15	(2) INFORMATION FOR SEQ ID NO:6:	
•	(i) SEQUENCE CHARACTERISTICS:	
20	(A) LENGTH: 1169 amino acids	
4	(B) TYPE: amino acid	
	(C) STRANDEDNESS: single	
25	(D) TOPOLOGY: unknown	
	(ii) MOLECULE TYPE: peptide	
ř.	•	
30	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:	
η	Gln Leu Ser Tyr Gly Tyr Asp Glu Lys Ser Thr Gly Gly Ile Ser Val	
35	1 5 10 15	
•	Pro Gly Pro Met Gly Pro Ser Gly Pro Arg Gly Leu Pro Gly Pro Pro	
40	20 25 30	
	Gly Ala Pro Gly Pro Gln Gly Phe Gln Gly Pro Pro Gly Glu Pro Gly	
	. 35 40 45	
45		
	Glu Pro Gly Ala Ser Gly Pro Met Gly Pro Arg Gly Pro Pro Gly Pro	
	50 55 60	
50		
	Pro Gly Lys Asn Gly Asp Asp Gly Glu Ala Gly Lys Pro Gly Arg Pro	
	65 70 75 80	

ı	Gly	Glu	Arg	Gly	Pro 85	Pro	Gly	Pro	Gln	Gly 90	Ala	Arg	Gly	Leu	Pro 95	Gly
5																
	Thr	Ala	Gly	Leu	Pro	Gly	Met	Lys	Gly	His	Arg	Gly	Phe	Ser	Gly	Leu
10				100					105					110		
				_	_,	_	_ •			•••	93	D	T	01 -		_
	Asp	GIÀ	115	Lys	GIÀ	Asp	Ala	120	Pro	Ala	GIY	PIO	125	Gly	GIU	Pro
15			113					120								
	Gly	Ser	Pro	Gly	Glu	Asn	Gly	Ala	Pro	Gly	Gln	Met	Gly	Pro	Arg	Gly
		130					135					140				
20																
	Leu	Pro	Gly	Glu	Arg	Gly	Arg	Pro	Gly	Ala		Gly	Pro	Ala	Gly	
	145					150					155					160
25) ra	Glv	Δan	Agn	Glv	Δla	Thr	Glv	Ala	Δla	Glv	Pro	Pro	Gly	Pro	Thr
•	πg	GIJ	ADII		165			O.F.		170	,			7	175	
•																
30	Gly	Pro	Ala	Gly	Pro	Pro	Gly	Phe	Pro	Gly	Ala	Val	Gly	Ala	Lys	Gly
				180					185					190		
25'			~ 3	D	a 1	a 1	D	3	01		61.	a]	Dw.e.	~ 1 ~	61	11-1
35	GIU	ATA	195	Pro	GIN	GIĀ	Pro	Arg 200	GIÀ	ser	GIU	GIY	205	Gln	GTÅ	vaı
Se.																
40	Arg	Gly	Glu	Pro	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Ala	Ala	Gly	Pro	Ala
		210					215					220				
45	-	Asn	Pro	Gly	Ala	_	Gly	Gln	Pro	Gly		Lys	Gly	Ala	Asn	Gly
	225					230					235					240
	Ala	Pro	Gly	Ile	Ala	Gly	Ala	Pro	Gly	Phe	Pro	Gly	Ala	Arg	Gly	Pro
50			•		245	-			-	250		-		,	255	
			•													

·	Ser	Gly	Pro	Gln	Gly	Pro	Gly	Gly	Pro	Pro	Gly	Pro	Lys	Gly	Asn	Ser
5				260					265					270		
	G1v	Glu	Pro	GJV	Δla	Pro	Gly	Ser	Two	Gly	n en	Thr	Glv	Λ Ι 5	Lys	C 1
	GLY	Giu	275	017	AIG	710	Gly	280	пув	GIY	voħ	1111	285	AIG	БÅЗ	GIY
10			2,,3					200					205			
•	Glu	Pro	Gly	Pro	Val	Gly	Val	Gln	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Glu
;		290					295					300				
15																
•	Glu	Gly	Lys	Arg	Gly	Ala	Arg	Gly	Glu	Pro	Gly	Pro	Thr	Gly	Leu	Pro
	305					310					315					320
20																
	Gly	Pro	Pro	Gly	Glu	Arg	Gly	Gly	Pro	Gly	Ser	Arg	Gly	Phe	Pro	Gly
					325					330					335	
25																
	Ala	qaA	Gly	Val	Ala	Gly	Pro	Lys	Gly	Pro	Ala	Gly	Glu	Arg	Gly	Ser
				340					345					350		
30	Pro	Gly	Pro	Ala	Gly	Pro	Lys	Gly	Ser	Pro	Gly	Glu	Ala	Gly	Arg	Pro
			355					360					365			
:																
35	Gly	Glu	Ala	Gly	Leu	Pro	Gly	Ala	Lys	Gly	Leu	Thr	Gly	Ser	Pro	Gly
•		370					375					380		•		
· • •																
40	Ser	Pro	Gly	Pro	Asp	Gly	Lys	Thr	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Gln
	385					390					395					400
45	Asp	Gly	Arg	Pro	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Ala	Arg	Gly	Gln	Ala
					405					410					415	
•																
50	Gly	Val	Met	Gly	Phe	Pro	Gly	Pro	Lys	Gly	Ala	Ala	Gly	Glu	Pro	Gly
50				420					425					430		

	Lys	Ala	-	Glu	Arg	Gly	Val		Gly	Pro	Pro	Gly		Val	Gly	Pro
5			435					440					445			
	Ala	Gly	Lys	Asp	Gly	Glu	Ala	Gly	Ala	Gln	Gly	Pro	Pro	Gly	Pro	Ala
10		450					455					460				
	Gly	Pro	Ala	Gly	Glu	Arg	Gly	Glu	Gln	Gly	Pro	Ala	Gly	Ser	Pro	Gly
	465					470					475					480
15	m 1	6 1-	g]	T 0	Dwo	61.	Dwa	71 -	C1.	Dva	Bro	alv	<i>c</i> 1	Ala	C1	T
	Pue	GIN	GIŞ	Leu	485	GIY	PIO	ALA		490	PIO	GLY	Giu	AIG	495	гув
20																
	Pro	Gly	Glu		Gly	Val	Pro	Gly	_	Leu	Gly	Ala	Pro	Gly	Pro	Ser
				500					505					510		
25	Gly	Ala	Arg	Gly	Glu	Arg	Gly	Phe	Pro	Gly	Glu	Arg	Gly	Val	Gln	Gly
			515					520					525			
30	Pro	Pro	Gly	Pro	Ala	Gly	Pro	Arg	Gly	Ala	Asn	Gly	Ala	Pro	Gly	Asn
		530					535					540				
		~3	*1-	•	01.	3		0 3		D	~1	210	Dwo	61	Com	61
35	Asp 545	GIŸ	Ala	ьув	GTĀ	550	Ala	grÅ	AIA	PIO	555	Ala	PIG	Gly	ser	560
40	Gly	Ala	Pro	Gly		Gln	Gly	Met	Pro		Glu	Arg	Gly	Ala		Gly
					565					570					575	
	Leu	Pro	Gly	Pro	Lys	Gly	Asp	Arg	Gly	Asp	Ala	Gly	Pro	Lys	Gly	Ala
				580					585					590		
	Asp	Gly	Ser	Pro	Gly	Lys	Asp	Gly	Val	Arg	Gly	Leu	Thr	Gly	Pro	Ile
50	-	-	595		-	-	.	600					605			
													•			

	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Ala	Pro	Gly	Asp	Lys	Gly	Glu	Ser	Gly
5		610			•		615					620				
	Pro	Ser	Gly	Pro	Ala	Gly	Pro	Thr	Gly	Ala	Arg	Gly	Ala	Pro	Gly	Asp
10	625					630					635					640
•																
	Arg	Gly	Glu	Pro	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Phe	Ala	Gly	Pro	Pro
					645					650					655	
15													_		_	
	Gly	Ala	Asp	-	Gln	Pro	Gly	Ala	-	Gly	Glu	Pro	Gly		Ala	Gly
				660					665					670		
20		_		_			_	_				-1	5		~ 3	
	Ala	Lys	-	Asp	ATA	GIA	Pro		GIĀ	Pro	Ala	GIY		Ala	Gly	Pro
			675					680					685			
25	D-10	~1. .	Dwa	T10	Cly	N cm	Va1	C1	מות	Pro	Gly	פות	Tare	Glv	Ala	7~~
	PIO	690	PIO	116	GIY	MBII	695	GIY	ALG	FIO	GIY	700	nya	GIŞ	ALG	ALY
		030					093					,,,,				
30	Glv	Ser	Ala	Glv	Pro	Pro	Glv	Ala	Thr	Gly	Phe	Pro	Gly	Ala	Ala	Gly
	705					710				-	715		•			720
•																
35'	Arg	Val	Gly	Pro	Pro	Gly	Pro	Ser	Gly	Asn	Ala	Gly	Pro	Pro	Gly	Pro
į					725					730					735	
10	Pro	Gly	Pro	Ala	Gly	Lys	Glu	Gly	Gly	Lys	Gly	Pro	Arg	Gly	Glu	Thr
				740					745					750		
4 5	Gly	Pro	Ala	Gly	Arg	Pro	Gly	Glu	Val	Gly	Pro	Pro	Gly	Pro	Pro	Gly
			755					760					765			
50	Pro			Glu	Lys	Gly			Gly	Ala	Asp			Ala	Gly	Ala
		770					775					780				

•	Pro	Gly	Thr	Pro	Gly	Pro	Gln	Gly	Ile	Ala	Gly	Gln	Arg	Gly	Val	Val
5	785					790				•	795					800
	Gly	Leu	Pro	Gly	Gln 805	Arg	Gly	Glu	Arg	Gly 810	Phe	Pro	Gly	Leu	Pro 815	Gly
10																
	Pro	Ser	Gly	G1u 820	Pro	Gly	Lys	Gln	G1y 825	Pro	Ser	GIÀ	Ala	Ser 830	Gly	Glu
15	Arg	Gly		Pro	Gly	Pro	Met		Pro	Pro	Gly	Leu		Gly	Pro	Pro
20			835					840					845			
	Gly	Glu 850	Ser	Gly	Arg	Glu	Gly 855	Ala	Pro	Ala	Ala	Glu 860	Gly	Ser	Pro	Gly
25	Arg	Asp	Gly	Ser	Pro	Gly	Ala	Lys	Gly	Asp	Arg	Gly	Glu	Thr	Gly	Pro
	865					870					875					880
30	Ala	Gly	Pro	Pro	Gly 885	Ala	Xaa	Gly	Ala	Xaa 890	Gly	Ala	Pro	Gly	Pro 895	Val
35	Gly	Pro	Ala	Gly 900	Lys	Ser	Gly	Asp	Arg 905	Gly	Glu	Thr	Gly	Pro 910	Ala	Gly
4 0	Pro	Ala	-	Pro	Val	Gly	Pro		Gly	Ala	Arg	Gly		Ala	Gly	Pro
			915					920					925			
4 5	Gln	Gly 930	Pro	Arg	Gly	Asp	Lys 935	_	Glu	Thr	Gly	Glu 940	Gln	Gly	Asp	Arg
50		Ile	Lys	Gly	His		Gly	Phe	Ser	Gly		Gln	Gly	Pro	Pro	
JU	945					950					955					960

_	Pro	Pro	Gly	Ser	Pro	Gly	Glu	Gln	Gly	Pro	Ser	Gly	Ala	Ser	Gly	Pro
5					965					970					975	
	Ala	Gly	Pro	Arg	Gly	Pro	Pro	Gly	Ser	Ala	Gly	Ala	Pro	Gly	Lys	gaA
10		-		980	•			•	985		-			990		
	Gly	Leu	Asn	Gly	Leu	Pro	Gly	Pro	Ile	Gly	Pro	Pro	Gly	Pro	Arg	Gly
15			995					1000)				1005	i		
	7~~	Thr	Glv	Δen	בומ	Glv	Pro	U = I	Gly	Pro	Pro	Glv.	Dro	Dro	Gly	Desa
	ALY	1010		nup	71.44	O.J	1015		Gry	110	110	1020		110	GIY	PIO
20			-										•			
	Pro	Gly	Pro	Pro	Gly	Pro	Pro	Ser	Ala	Gly	Phe	Asp	Phe	Ser	Phe	Leu
	1025	5				1030)				1035	;				1040
25																
6.	Pro	Gln	Pro	Pro	Gln	Glu	Lys	Ala	His	Asp	Gly	Gly	Arg	Tyr	Tyr	Arg
					1045	5				1050)				1055	5
30		_	_		_		_		_		_	_	_			
	Ala	Arg	ser			Ala	Arg	Lys	Lys 1065		Lys	Asn	Cys	_	Arg	His
<i>.</i>				3000						>					3	
				1060)				100.					1070	•	
<i>;</i> 35	Ser	Leu	Tvr			Phe	Ser	Asp		Glv	Tro	Asn	Asp			Val
	Ser	Leu	Tyr 1079	Val		Phe	Ser	Asp	Val	Gly	Trp	Asn	Asp 1085	Trp	Ile	Val
35	Ser	Leu		Val		Phe	Ser	_	Val	Gly	Trp	Asn	_	Trp		Val
			1075	Val	Asp			1080	Val				1085	Trp		
35			1075 Pro	Val	Asp			1080	Val				1085 Asp	Trp	Ile	
35° 40		Pro	1075 Pro	Val	Asp		Ala	1080	Val			Gly	1085 Asp	Trp	Ile	
35	Ala	Pro 1090	1075 Pro	Val Gly	Азр	Gln	Ala 1095	1080	Val) Tyr	Сув	His	Gly 1100	1085 Asp	Trp Cys	Ile	Phe
35° 40	Ala	Pro 1090 Leu	1075 Pro	Val Gly	Азр	Gln	Ala 1095 Asn	1080	Val) Tyr	Сув	His	Gly 1100 Ala	1085 Asp	Trp Cys	Ile Pro	Phe
35 40 45	Pro	Pro 1090 Leu	Pro	Val Gly Asp	Asp Tyr His	Gln Leu 1110	Ala 1095 Asn	Phe Ser	Val Tyr Thr	Cys	His His	Gly 1100 Ala	Asp	Trp Cys Val	Ile Pro	Phe Thr 1120
35° 40	Pro	Pro 1090 Leu	Pro	Val Gly Asp	Asp Tyr His	Gln Leu 1110	Ala 1095 Asn	Phe Ser	Val Tyr Thr	Cys Asn Pro	His His 1115	Gly 1100 Ala	Asp	Trp Cys Val	Ile Pro Gln Val	Phe Thr 1120 Pro
35 40 45	Pro	Pro 1090 Leu	Pro	Val Gly Asp	Asp Tyr His	Gln Leu 1110	Ala 1095 Asn	Phe Ser	Val Tyr Thr	Cys	His His 1115	Gly 1100 Ala	Asp	Trp Cys Val	Ile Pro	Phe Thr 1120 Pro

	Thr	Glu	Leu	Ser	Ala	Ile	Ser	Met	Leu	Tyr	Leu	Asp	Glu	Tyr	Asp	Lys	
5				1140)				1149	5				115	D		
	Val	Val	Leu	Lys	Asn	Tyr	Gln	Glu	Met	Val	Val	Glu	Gly	Cys	Gly	Сув	
,· 10			1155	5				1160	0				116	5			
	Arg																
\$																	
15	(2) INFO	RMAT]	ION I	FOR S	SEQ I	ID NO	0:7:										
20	(i)	SEQU	JENCI	S CH	RACT	reri:	STIC	S:									
				NGTH:			_	pairs	3								•
				PE: r RANDE				l e									
25	•			POLOG													
30	(ii)	MOLI	ECULI	TYI	PE: o	DNA											
	(xi)	SEQU	JENCI	DES	CRI	PTIO	1: SI	EQ II	O NO:	7:							
35°.	GGGAAGGA'	TT TO	CATT	rtccc	AGC	CTGT	CTTA	TGG	CTATO	AT (gagai	ATC#	AA C	CGGA	GAAT	ŗ `	60
Ţ:	TTCCGTGC	CT GO	3CCC(CATGO	GTO	CCT	CTGG	TCC	rcgro	GT (CTCC	CTGGC	c c	CCT	GTG	2	120
40	ACCTGGTC	cc cz	AAGG(TTCC	: AAG	GTC	ccc	TGG	rgago	CCT (3GCG1	AGCCI	rg gj	AGCT	CAGO	}	180
45	TCCCATGG	ST CO	CCG	AGGTO	ccc	CAG	FTCC	ccc	rggaj	AAG J	AATGO	AGAT	G A	rggg(BAAGO	:	240
	TGGAAAAC	CT GO	STCGT	CCT	GTO	BAGC	TGG	GCC	CCT	GG (CCTC	AGGG1	rg cr	rcgao	GATT	r	300
50	GCCCGGAAG	CA GO	CTGGC	CTC	сто	GAA1	rgaa	GGG?	ACACA	\GA (GGTT:	CAGI	rg gr	rttg(ATGO	3	360
1	TGCCAAGG	GA GA	ATGCI	rggto	CTC	CTG	STCC	TAAC	GGT	GAG (CCTGO	CAGO	cc ca	rggto	BAAA		420

480	GCCCTGGAGC	GAGAGAGGTC	CCTGCCTGGT	GCCCCCGTGG	GGTCAGATGG	TGGAGCTCCT	
540	GGCCCCCTGG	GGTGCTGCCG	TGGTGCTAÇT	GTGGAAATGA	GCTGGTGCTC	CCCTGGCCCT	
600	AGGGTGAAGC	GTTGGTGCTA	CCCTGGTGCT	CTCCTGGCTT	CCCGCTGGTC	TCCCACCGGC	
660	AGCCTGGCCC	GTGCGTGGTG	TCCCCAGGGT	GCTCTGAAGG	GGGCCCCGAG	TGGTCCCCAA	
720	GACAGCCTGG	GGTGCTGATG	TGGAAACCCT	CTGGCCCTGC	GCTGGTGCTG	CCCTGGCCCT	
780	CTGGTGCCCG	CCTGGCTTCC	TGCTGGTGCT	CTCCTGGTAT	GCCAATGGTG	TGCTAAAGGT	
·8 4 0	ACAGCGGTGA	CCCAAGGGTA	CCCTCCTGGT	GCCCGGCGG	GGACCCCAGG	AGGCCCCTCT	
900	GCCCTGTTGG	GGAGAGCCTG	TGGTGCTAAG	AAGGAGACAC	CCTGGCAGCA	ACCTGGTGCT	
960	GAGGTGAACC	CGAGGAGCTC	GGAAGGAAAG	CTGCTGGAGA	CCCCCTGGCC	TGTTCAAGGA	
1020	GCCGTGGTTT	GGACCTGGTA	CGAGCGTGGT	GACCCCCTGG	GGCCTGCCCG	CGGACCCACT	
1080	GTTCTCCTGG	GGTGAACGTG	GGGTCCCGCT	CTGGTCCCAA	GATGGTGTTG	CCCTGGCGCA	
1140	CTGGTCTGCC	CCCGGTGAAG	AGCTGGTCGT	CTCCTGGTGA	CCCAAAGGAT	CCCCGCTGGC	
1200	AAACTGGCCC	CCTGATGGCA	CAGCCCTGGT	GAAGCCCTGG	GGTCTGACTG	TGGTGCCAAG	
1260	GTGCCCGTGG	GGCCCACCTG	CGGACCCCCA	ATGGTCGCCC	GCCGGTCAAG	CCCTGGTCCC	
1320	CCGGCAAGGC	GCTGGAGAGC	TAAAGGTGCT	TCCCTGGACC	GTGATGGGAT	TCAGGCTGGT	
1380	AAGATGGAGA	CCTGCTGGCA	CGCTGTCGGT	GACCCCCTGG	GGTGTTCCCG	TGGAGAGCGA	
1440	GTGAACAAGG	GGCGAGAGAG	TGGTCCCGCT	CTGGCCCTGC	CAGGGACCCC	GGCTGGAGCT	

·;;

CCCTGCTGGC	TCCCCCGGAT	TCCAGGGTCT	CCCTGGTCCT	GCTGGTCCTC	CAGGTGAAGC	1500
AGGCAAACCT	GGTGAACAGG	GTGTTCCTGG	AGACCTTGĢC	GCCCCTGGCC	CCTCTGGAGC	1560
AAGAGGCGAG	AGAGGTTTCC	CTGGCGAGCG	TGGTGTGCAA	GGTCCCCCTG	GTCCTGCTGG	1620
ACCCCGAGGG	GCCAACGGTG	CTCCCGGCAA	CGATGGTGCT	AAGGGTGATG	CTGGTGCCCC	1680
TGGAGCTCCC	GGTAGCCAGG	GCGCCCTGG	CCTTCAGGGA	ATGCCTGGTG	AACGTGGTGC	1740
AGCTGGTCTT	CCAGGGCCTA	AGGGTGACAG	AGGTGATGCT	GGTCCCAAAG	GTGCTGATGG	1800
CTCTCCTGGC	AAAGATGGCG	TCCGTGGTCT	GACCGGCCCC	ATTGGTCCTC	CTGGCCCTGC	1860
TGGTGCCCCT	GGTGACAAGG	GTGAAAGTGG	TCCCAGCGGC	CCTGCTGGTC	CCACTGGAGC	1920
TCGTGGTGCC	CCCGGAGACC	GTGGTGAGCC	TGGTCCCCCC	GGCCCTGCTG	GCTTTGCTGG	1980
CCCCCTGGT	GCTGACGGCC	AACCTGGTGC	TAAAGGCGAA	CCTGGTGATG	CTGGTGCCAA	2040
AGGCGATGGG	TCCCCTGGG	CCTGCCGGAC	CCGCTGGACC	CCCTGGCCCC	ATTGGTAATG	2100
TTGGTGCTCC	TGGAGCCAAA	GGTGCTCGCG	GCAGCGCTGG	TCCCCCTGGT	GCTACTGGTT	2160
TCCCTGGTGC	TGCTGGCCGA	GTCGGTCCTC	CTGGCCCCTC	TGGAAATGCT	GGACCCCCTG	2220
GCCCTCCTGG	TCCTGCTGGC	AAAGAAGGCG	GCAAAGGTCC	CCGTGGTGAG	ACTGGCCCTG	2280
CTGGACGTCC	TGGTGAAGTT	GGTCCCCCTG	GTCCCCCTGG	CCCTGCTGGC	GAGAAAGGAT	2340
CCCCTGGTGC	TGATGGTCCT	GCTGGTGCTC	CTGGTACTCC	CGGGCCTCAA	GGTATTGCTG	2400
GACAGCGTGG	TGTGGTCGGC	CTGCCTGGTC	AGAGAGGAGA	GAGAGGCTTC	CCTGGTCTTC	2460

	Credececte	IGGIGAACCI	GGCAAACAAG	GICCCICIGG	AGCAAGIGGT	GAACGIGGIC	2520
	CCCCCGGTCC	CATGGGCCCC	CCTGGATTGG	CTGGACCCCC	TGGTGAATCT	GGACGTGAGG	2580
10	GGGCTCCTGC	TGCCGAAGGT	TCCCCTGGAC	GAGACGGTTC	TCCTGGCGCC	AAGGGTGACC	2640
P.,	GTGGTGAGAC	CGGCCCCGCT	GGACCCCCTG	GTGCTCTGGT	GCTCTGGTGC	CCCTGGCCCC	2700
15	GTTGGCCCTG	CTGGCAAGAG	TGGTGATCGT	GGTGAGACTG	GTCCTGCTGG	TCCCGCCGGT	2760
	CCCGTCGGCC	CCGCTGGCGC	CCGTGGCCCC	GCCGGACCCC	AAGGCCCCCG	TGGTGACAAG	2820
20	GGTGAGACAG	GCGAACAGGG	CGACAGAGGC	ATAAAGGGTC	ACCGTGGCTT	CTCTGGCCTC	2880
25	CAGGGTCCCC	CTGGCCCTCC	TGGCTCTCCT	GGTGAACAAG	GTCCCTCTGG	AGCCTCTGGT	2940
25	CCTGCTGGTC	CCCGAGGTCC	CCCTGGCTCT	GCTGGTGCTC	CTGGCAAAGA	TGGACTCAAC	3000
30	GGTCTCCCTG	GCCCCATTGG	GCCCCTGGT	CCTCGCGGTC	GCACTGGTGA	TGCTGGTCCT	3060
3	GTTGGTCCCC	CCGGCCCTCC	TGGACCTCCT	GGTCCCCCTG	GTCCTCCCAG	CGCTGGTTTC	3120
35 ⁻	GACTTCAGCT	TCCTCCCCCA	GCCACCTCAA	GAGAAGGCTC	ACGATGGTGG	CCGCTACTAC	3180
.2.	CGGGCTAGAT	CCCAGCGGGC	CAGGAAGAAG	AATAAGAACT	GCCGGCGCCA	CTCGCTCTAT	3240
40	GTGGACTTCA	GCGATGTGGG	CTGGAATGAC	TGGATTGTGG	CCCCACCAGG	CTACCAGGCC	3300
45	TTCTACTGCC	ATGGGGACTG	CCCCTTTCCA	CTGGCTGACC	ACCTCAACTC	AACCAACCAT	3360
	GCCATTGTGC	AGACCCTGGT	CAATTCTGTC	AATTCCAGTA	TCCCCAAAGC	CTGTTGTGTG	3420
50	CCCACTGAAC	. TGAGTGCCAT	CTCCATGCTG	TACCTGGATG	AGTATGATAA	GGTGGTACTG	3480

5	AAAAATTATC AGGAGATGGT AGTAGAGGGA TGTGGGTGCC GCTAAAAGCT T	3531
	(2) INFORMATION FOR SEQ ID NO:8:	
-10	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1171 amino acids (B) TYPE: amino acid	
∵15	(C) STRANDEDNESS: single (D) TOPOLOGY: unknown	,
20	(ii) MOLECULE TYPE: peptide	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:	
25	Gln Leu Ser Tyr Gly Tyr Asp Glu Lys Ser Thr Gly Gly Ile Ser v	Val
30	Pro Gly Pro Met Gly Pro Ser Gly Pro Arg Gly Leu Pro Gly Pro 1	Pro
35	Gly Ala Pro Gly Pro Gln Gly Phe Gln Gly Pro Pro Gly Glu Pro G	31y
40	Glu Pro Gly Ala Ser Gly Pro Met Gly Pro Arg Gly Pro Pro Gly 1	Pro
· 4 5	Pro Gly Lys Asn Gly Asp Asp Gly Glu Ala Gly Lys Pro Gly Arg	Pro 80
50	Gly Glu Arg Gly Pro Pro Gly Pro Gln Gly Ala Arg Gly Leu Pro 6	31y
55	Thr Ala Gly Leu Pro Gly Met Lys Gly His Arg Gly Phe Ser Gly 100 105 110	Leu

	Asp	Gly	Ala	Lys	Gly	Asp	Ala	Gly	Pro	Ala	Gly	Pro	Lys	Gly	Glu	Pro
			115					120					125			
5										•						
	Gly	Ser	Pro	Gly	Glu	Asn	Gly	Ala	Pro	Gly	Gln	Met	Gly	Pro	Arg	Gly
		130					135					140				
10																
	Leu	Pro	Gly	Glu	Arg	Gly	Arg	Pro	Gly	Ala	Pro	Gly	Pro	Ala	Gly	Ala
	145					150					155					160
15																
	Arg	Gly	Asn	qaA		Ala	Thr	Gly	Ala	Ala	Gly	Pro	Pro	Gly	Pro	Thr
					165					170					175	
20										_						
	Gly	Pro	Ala	-	Pro	Pro	Gly	Phe		Gly	Ala	Val	Gly		Lys	Gly
				180					185					190		
25			~ 3		0 1	a 1	D	.	63	g.,,	01	~1	Dwa	a1 -	61	17. 1
	Glu	Ala	_	Pro	GIN	GIÀ	Pro	_	GIY	ser	GIU	GIĀ		GIN	GIĀ	Val
•			195					200					205			
30	7~~	Clv.	Glu	Pro	G1 v	Dro	Pro	Glv	Dro	בומ	Glv	Δla	Δla	Glv	Pro	Ala
	Arg	210	GIU	210		210	215	O.J		7624	02]	220		,		
		210														
35	Glv	Asn	Pro	Glv	Ala	Asp	Glv	Gln	Pro	Glv	Ala	Lys	Gly	Ala	Asn	Gly
33	225		•	,		230	,			•	235	•	•			240
	Ala	Pro	Gly	Ile	Ala	Gly	Ala	Pro	Gly	Phe	Pro	Gly	Ala	Arg	Gly	Pro
40					245					250					255	
	Ser	Gly	Pro	Gln	Gly	Pro	Gly	Gly	Pro	Pro	Gly	Pro	Lys	Gly	Asn	Ser
45				260					265					270		
	Gly	Glu	Pro	Gly	Ala	Pro	Gly	Ser	Lys	Gly	Asp	Thr	Gly	Ala	Lys	Gly
50			275					280					285			

•	Glu	Pro	Gly	Pro	Val	Gly	Val	Gln	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Glu
5		290					295					300				
u.	Glu	Gly	Lys	Arg	Gly	Ala	Arg	Gly	Glu	Pro	Gly	Pro	Thr	Gly	Leu	Pro
10	305					310					315					320
	Gly	Pro	Pro	Gly		Arg	Gly	Gly	Pro		Ser	Arg	Gly	Phe	Pro	Gly
15					325					330					335	
	Ala	Asp	Gly		Ala	Gly	Pro	Lys	_	Pro	Ala	Gly	Glu		Gly	Ser
20				340					345					350		
	Pro	Gly	Pro 355	Ala	Gly	Pro	Lys	Gly 360	Ser	Pro	Gly	Glu	Ala 365	Gly	Arg	Pro
25			333					360					363			
	Gly	Glu 370	Ala	Gly	Leu	Pro	Gly 375	Ala	Lys	Gly	Leu	Thr 380	Gly	Ser	Pro	Gly
30		370					3,3					300				
	Ser	Pro	Gly	Pro	Asp	Gly	Lys	Thr	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Gln
∆ 35°	385					390				-	395					400
	Asp	Gly	Arg	Pro	Gly 405	Pro	Pro	Gly	Pro	Pro	Gly	Ala	Arg	Gly	Gln 415	Ala
:																
40	Gly	Val	Met	Gly	Phe	Pro	Gly	Pro	Lys	Gly	Ala	Ala	Gly	Glu	Pro	Gly
				420	,				425					430		
45	Lys	Ala	Gly	Glu	Arg	Gly	Val	Pro	Gly	Pro	Pro	Gly	Ala	Val	Gly	Pro
•			435					440					445			
50	Ala	_	_	Asp	Gly	Glu		•	Ala	Gln	Gly		Pro	Gly	Pro	Ala
		450					455					460				
•																

	Gly	Pro	Ala	Gly	Glu	Arg	Gly	Glu	Gln	Gly	Pro	Ala	Gly	Ser	Pro	Gly
5	465					470					475					480
	Phe	Gln	Gly	Leu	Pro	Gly	Pro	Ala	Gly	Pro	Pro	Gly	Glu	Ala	Gly	Lys
10					485					490					495	
	Pro	Gly	Glu	Gln	Gly	Val	Pro	Gly	Asp	Leu	Gly	Ala	Pro	Gly	Pro	Ser
15				500					505					510		
15																
	Gly	Ala	Arg	Gly	Glu	Arg	Gly	Phe	Pro	Gly	Glu	Arg	Gly	Val	Gln	Gly
			515		•			520					525			
20										_						
•	Pro		Gly	Pro	Ala	Gly		Arg	Gly	Ala	Asn		Ala	Pro	Gly	Asn
		530					535					540				
25								-	_ =	_			_		_	
٠.,	_	Gly	Ala	Lys	Gly		Ala	Gly	Ala	Pro		Ala	Pro	GIÀ	Ser	Gln
:	545					550					55 5					560
30			_		•	63	a1		D	41	a1	3	~1	21-	23-	01
	Gly	Ala	Pro	GIĀ			GIĀ	met	Pro		GIU	Arg	GTĀ	ATA		Gly
					565					570					575	
35	T -41	D=0	C1.	Dro	Tara	GI.v	A co	n.c.	G] v) an	פוג	G) v	Pro	Tare	Glv	Ala
•	Leu	PIO	GIŞ	580	пåз	GIY	veh	ALG	585	vob	A.L	GLY		590	GIJ	nia
				300					505							
40	Aen	ผาง	Ser	Pro	Glv	Lvs	Asp	Glv	Val	Arg	Glv	Leu	Thr	Glv	Pro	Ile
••	-mp	u_y	595		1	-,-	· <u>-</u> -	600		3	2		605			
÷			333					•••								
:_	Glv	Pro	Pro	Gly	Pro	Ala	Gly	Ala	Pro	Gly	Asp	Lys	Gly	Glu	Ser	Gly
45	2	610		•			615			•	-	620	_			·
•																
	Pro	Ser	Gly	Pro	Ala	Gly	Pro	Thr	Gly	Ala	Arg	Gly	Ala	Pro	Gly	Asp
50	625		•			630			-		635					640

	Arg	Gly	Glu	Pro	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Phe	Ala	Gly	Pro	Pro
5					645					650					655	
										•						
	Gly	Ala	Asp	Gly	Gln	Pro	Gly	Ala	Lys	Gly	Glu	Pro	Gly	Asp	Ala	Gly
				660					665					670		
10																
	Ala	Lys	Gly	Asp	Ala	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Pro	Ala	Gly	Pro
			675					680					685			
15																
•	Pro	Gly	Pro	Ile	Gly	Asn	Val	Gly	Ala	Pro	Gly	Ala	Lys	Gly	Ala	Arg
		690					695					700				
20	•															
		Ser	Ala	Gly	Pro		Gly	Ala	Thr	Gly		Pro	Gly	Ala	Ala	Gly
	705					710					715					720
25																
	Arg	Val	Gly	Pro		Gly	Pro	Ser	Gly		Ala	Gly	Pro	Pro	_	Pro
					725					730					735	
; 30	_					_				_		_	_		_	
	Pro	GIY	Pro		GIÀ	гуs	GIu	GIÀ		Lys	GIY	Pro	Arg		Glu	Thr
				740					745					750		
	~1	Dwa	71 n	~1	N	Dwa	a 1	~1	171	63	D	7	01	D	D	Gly
35	GIY	PIO	755	GIY	Arg	PIO	GIY	760	Val	GIÅ	PIO	PIO	765	PIO	PLO	GIĀ
			,,,,					700					103			
:	Pro	Ala	Gly	Glu	Lvs	Glv	Ser	Pro	Glv	Ala	Asp	ตาง	Pro	Ala	G] v	Δla
40		770			-3-	1	775		- -,			780		••••	 3	
	Pro	Gly	Thr	Pro	Gly	Pro	Gln	Gly	Île	Ala	Gly	Gln	Arg	Gly	Val	Val
45	785	-			-	790		•			795		_	•		800
	Gly	Leu	Pro	Gly	Gln	Arg	Gly	Glu	Arg	Gly	Phe	Pro	Gly	Leu	Pro	Gly
50					805					810			-		815	-
· ·																

55[°]

	Pro	Ser	Gly	Glu	Pro	Gly	Lys	Gln	Gly	Pro	Ser	Gly	Ala	Ser	Gly	Glu
5				820					825					830		
	Arg	Gly	Pro	Pro	Gly	Pro	Met	Gly	Pro	Pro	Gly	Leu	Ala	Gly	Pro	Pro
			835					840					845			
10																
	Gly	Glu	Ser	Gly	Arg	Glu	Gly	Ala	Pro	Gly	Ala	Glu	Gly	Ser	Pro	Gly
		850					855					860				
15																
	Arg	Asp	Gly	Ser	Pro	Gly	Ala	Lys	Gly	Asp	Arg	Gly	Glu	Thr	Gly	Pro
	865					870					875					880
20									_				_			
	Ala	Gly	Pro	Pro	-	Ala	Pro	Gly	Ala		Gly	Ala	Pro	Gly		Val
					885					890					895	
25		_	- -		_	_		_	_	~ 3	~ 1		63	D	-1-	a 1
	Gly	Pro	Ala	-	гåа	ser	GTÅ	Asp		GIY	GIU	Thr	GIÀ	Pro	Ala	GIÅ
				900					905					910		
30	D-1-	27	~1	Dro	wal.	Gly	Dro	λla	Gly	בות	Ara	Gly	Pro	Ala	Glv	Pro
	PIO	ATA	915	PIO	Val	GLY	710	920	GIY	ALU	,g	o _z ,	925		O.J	
			313					720	•				,,,,			
35	Gln	Glv	Pro	Ara	Glv	Asp	Lvs	Glv	Glu	Thr	Glv	Glu	Gln	Gly	Asp	Arq
	· · · ·	930			-		935	1			2	940				
40:	Gly	Ile	Lys	Gly	His	Arg	Gly	Phe	Ser	Gly	Leu	Gln	Gly	Pro	Pro	Gly
40 ⁻	945		-			950					955					960
	Pro	Pro	Gly	Ser	Pro	Gly	Glu	Gln	Gly	Pro	Ser	Gly	Ala	Ser	Gly	Pro
45					965					970					975	
	Ala	Gly	Pro	Arg	Gly	Pro	Pro	Gly	Ser	Ala	Gly	Ala	Pro	Gly	Lys	Asp
50				980					985					990		

	Gly	Leu	Asn	Gly	Leu	Pro	Gly	Pro	Ile	Gly	Pro	Pro	Gly	Pro	Arg	Gly
5			995					1000)				100	5		
	Arg	Thr	Gly	Asp	Ala	Gly	Pro	Val	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Pro
10		1010)				1015	5				1020)			
	Pro	Gly	Pro	Pro	Gly	Pro	Pro	Ser	Ala	Gly	Phe	Asp	Phe	Ser	Phe	Leu
15	1025	5				1030)				1035	5				1040
	Pro	Gln	Pro	Pro	Gln	Glu	Lys	Ala	His	qaA	Gly	Gly	Arg	Tyr	Tyr	Arg
20					1045	5				1050)				1055	5
20	7.1 o	n-a	Ser	בומ	T.e.u) an	The	7.55	The same	Caro	Dho	50×	Co~	The	G1	T
	AIA	Arg	Ser	1060		App	1111	ABII	1065		PHE	Ser	Ser	1070		гув
25	_	_	_		_		_	_				_	_	_		
	Asn	Cys	Cys 1079		Arg	GIN	Leu	Tyr 1080		Asp	Phe	Arg	1089	-	Leu	Gly
			1071	-				100	•	•			100.	•		
30	Trp	Lys	Trp	Ile	His	Glu	Pro	Lys	Gly	Tyr	His	Ala	Asn	Phe	Cys	Leu
		1090	ס				1099	5				1100)			
35						_										
	Gly 1105		Cys	Pro	Tyr	Ile 1110		Ser	Leu	Asp			Tyr	Ser	Lys	
	110	,				111(•				1115	,				1120
40	Leu	Ala	Leu	Tyr	Asn	Gln	His	Asn	Pro	Gly	Ala	Ser	Ala	Ala	Pro	Сув
					1125	5				1130)				1135	5
		_		_	_											
45	Cys	Val	Pro			Leu	Glu	Pro			Ile	Val	Tyr	_		Gly
				1140	,				1145	•				1150)	
50	Arg	Lys	Pro	Lys	Val	Glu	Gln	Leu	Ser	Asn	Met	Ile	Val	Arg	Ser	Сув
			1155	5				1160)				1165	5		

Lys	Сув	Sea
	1170)

5

10

15

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3541 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

20

25

35

40

45

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GGGAAGGATT TCCATTTCCC AGCTGTCTTA TGGCTATGAT GAGAAATCAA CCGGAGGAAT 60 TTCCGTGCCT GGCCCCATGG GTCCCTCTGG TCCTCGTGGT CTCCCTGGCC CCCCTGGTGC ACCTGGTCCC CAAGGCTTCC AAGGTCCCCC TGGTGAGCCT GGCGAGCCTG GAGCTTCAGG TCCCATGGGT CCCCGAGGTC CCCCAGGTCC CCCTGGAAAG AATGGAGATG ATGGGGAAGC 240 TGGAAAACCT GGTCGTCCTG GTGAGCGTGG GCCTCCTGGG CCTCAGGGTG CTCGAGGATT 300 GCCCGGAACA GCTGGCCTCC CTGGAATGAA GGGACACAGA GGTTTCAGTG GTTTGGATGG 360 TGCCAAGGA GATGCTGGTC CTGCTGGTCC TAAGGGTGAG CCTGGCAGCC CTGGTGAAAA 420 TGGAGCTCCT GGTCAGATGG GCCCCCGTGG CCTGCCTGGT GAGAGAGGTC GCCCTGGAGC 480 CCCTGGCCCT GCTGGTGCTC GTGGAAATGA TGGTGCTACT GGTGCTGCCG GGCCCCCTGG 540 TCCCACCGGC CCCGCTGGTC CTCCTGGCTT CCCTGGTGCT GTTGGTGCTA AGGGTGAAGC 600

;

TGGTCCCCAA	GGGCCCCGAG	GCTCTGAAGG	TCCCCAGGGT	GTGCGTGGTG	AGCCTGGCCC	660
CCCTGGCCCT	GCTGGTGCTG	CTGGCCCTGC	TGGAAACCCT	GGTGCTGATG	GACAGCCTGG	720
TGCTAAAGGT	GCCAATGGTG	CTCCTGGTAT	TGCTGGTGCT	CCTGGCTTCC	CTGGTGCCCG	780
AGGCCCCTCT	GGACCCCAGG	GCCCCGGCGG	CCCTCCTGGT	CCCAAGGGTA	ACAGCGGTGA	840
ACCTGGTGCT	CCTGGCAGCA	AAGGAGACAC	TGGTGCTAAG	GGAGAGCCTG	GCCCTGTTGG	900
TGTTCAAGGA	CCCCCTGGCC	CTGCTGGAGA	GGAAGGAAAG	CGAGGAGCTC	GAGGTGAACC	960
CGGACCCACT	GGCCTGCCCG	GACCCCTGG	CGAGCGTGGT	GGACCTGGTA	GCCGTGGTTT	1020
CCCTGGCGCA	GATGGTGTTG	CTGGTCCCAA	GGGTCCCGCT	GGTGAACGTG	GTTCTCCTGG	1080
CCCCGCTGGC	CCCAAAGGAT	CTCCTGGTGA	AGCTGGTCGT	CCCGGTGAAG	CTGGTCTGCC	1140
TGGTGCCAAG	GGTCTGACTG	GAAGCCCTGG	CAGCCCTGGT	CCTGATGGCA	AAACTGGCCC	1200
CCCTGGTCCC	GCCGGTCAAG	ATGGTCGCCC	CGGACCCCCA	GGCCCACCTG	GTGCCCGTGG	1260
TCAGGCTGGT	GTGATGGGAT	TCCCTGGACC	TAAAGGTGCT	GCTGGAGAGC	CCGGCAAGGC	1320
TGGAGAGCGA	GGTGTTCCCG	GACCCCCTGG	CGCTGTCGGT	CCTGCTGGCA	AAGATGGAGA	1380
GGCTGGAGCT	CAGGGACCCC	CTGGCCCTGC	TGGTCCCGCT	GGCGAGAGAG	GTGAACAAGG	1440
CCCTGCTGGC	TCCCCCGGAT	TCCAGGGTCT	CCCTGGTCCT	GCTGGTCCTC	CAGGTGAAGC	1500
AGGCAAACCT	GGTGAACAGG	GTGTTCCTGG	AGACCTTGGC	GCCCCTGGCC	CCTCTGGAGC	1560
AAGAGGCGAG	AGAGGTTTCC	CTGGCGAGCG	TGGTGTGCAA	GGTCCCCCTG	GTCCTGCTGG	1620

ACCCCGAGGG	GCCAACGGTG	CTCCCGGCAA	CGATGGTGCT	AAGGGTGATG	CTGGTGCCCC	1680
TGGAGCTCCC	GGTAGCCAGG	GCGCCCTGG	CCTTCAGGĢA	ATGCCTGGTG	AACGTGGTGC	1740
AGCTGGTCTT	CCAGGGCCTA	AGGGTGACAG	AGGTGATGCT	GGTCCCAAAG	GTGCTGATGG	1800
CTCTCCTGGC	AAAGATGGCG	TCCGTGGTCT	GACCGGCCCC	ATTGGTCCTC	CTGGCCCTGC	1860
TGGTGCCCCT	GGTGACAAGG	GTGAAAGTGG	TCCCAGCGGC	CCTGCTGGTC	CCACTGGAGC	1920
TCGTGGTGCC	CCCGGAGACC	GTGGTGAGCC	TGGTCCCCC	GGCCCTGCTG	GCTTTGCTGG	1980
CCCCCTGGT	GCTGACGGCC	AACCTGGTGC	TAAAGGCGAA	CCTGGTGATG	CTGGTGCCAA	2040
AGGCGATGCT	GGTCCCCCTG	GGCCTGCCGG	ACCCGCTGGA	CCCCCTGGCC	CCATTGGTAA	2100
TGTTGGTGCT	CCTGGAGCCA	AAGGTGCTCG	CGGCAGCGCT	GGTCCCCCTG	GTGCTACTGG	2160
TTTCCCTGGT	GCTGCTGGCC	GAGTCGGTCC	TCCTGGCCCC	TCTGGAAATG	CTGGACCCCC	2220
TGGCCCTCCT	GGTCCTGCTG	GCAAAGAAGG	CGGCAAAGGT	CCCCGTGGTG	AGACTGGCCC	2280
TGCTGGACGT	CCTGGTGAAG	TTGGTCCCCC	TGGTCCCCCT	GGCCCTGCTG	GCGAGAAAGG	2340
ATCCCCTGGT	GCTGATGGTC	CTGCTGGTGC	TCCTGGTACT	CCCGGGCCTC	AAGGTATTGC	2400
TGGACAGCGT	GGTGTGGTCG	GCCTGCCTGG	TCAGAGAGGA	GAGAGAGGCT	TCCCTGGTCT	2460
TCCTGGCCCC	TCTGGTGAAC	CTGGCAAACA	AGGTCCCTCT	GGAGCAAGTG	GTGAACGTGG	2520
TCCCCCCGGT	CCCATGGGCC	CCCCTGGATT	GGCTGGACCC	CCTGGTGAAT	CTGGACGTGA	2580
GGGGGCTCCT	GCTGCCGAAG	GTTCCCCTGG	ACGAGACGGT	TCTCCTGGCG	CCAAGGGTGA	2640

. 20

i

50-

CCGTGGTGAG	ACCGGCCCCG	CTGGACCCCC	TGGTGCTCCT	GGTGCTCCTG	GTGCCCCTGG	2700
CCCCGTTGGC	CCTGCTGGCA	AGAGTGGTGA	TCGTGGTGAG	ACTGGTCCTG	CTGGTCCCGC	2760
CGGTCCCGTC	GGCCCCGCTG	GCGCCCGTGG	CCCCGCCGGA	CCCCAAGGCC	CCCGTGGTGA	2820
CAAGGGTGAG	ACAGGCGAAC	AGGGCGACAG	AGGCATAAAG	GGTCACCGTG	GCTTCTCTGG	2880
CCTCCAGGGT	CCCCTGGCC	CTCCTGGCTC	TCCTGGTGAA	CAAGGTCCCT	CTGGAGCCTC	2940
TGGTCCTGCT	GGTCCCCGAG	GTCCCCCTGG	CTCTGCTGGT	GCTCCTGGCA	AAGATGGACT	3000
CAACGGTCTC	CCTGGCCCCA	TTGGGCCCCC	TGGTCCTCGC	GGTCGCACTG	GTGATGCTGG	3060
TCCTGTTGGT	CCCCCCGGCC	CTCCTGGACC	TCCTGGTCCC	CCTGGTCCTC	CCAGCGCTGG	3120
TTTCGACTTC	AGCTTCCTCC	CCCAGCCACC	TCAAGAGAAG	GCTCACGATG	GTGGCCGCTA	3180
CTACCGGGCT	AGATCTGCCC	TGGACACCAA	CTATTGCTTC	AGCTCCACGG	AGAAGAACTG	3240
CTGCGTGCGG	CAGCTGTACA	TTGACTTCCG	CAAGGACCTC	GGCTGGAAGT	GGATCCACGA	3300
GCCCAAGGGC	TACCATGCCA	ACTTCTGCCT	CGGGCCCTGC	CCCTACATTT	GGAGCCTGGA	3360
CACGCAGTAC	AGCAAGGTCC	TGGCCCTGTA	CAACCAGCAT	AACCCGGGCG	CCTCGGCGGC	3420
GCCGTGCTGC	GTGCCGCAGG	CGCTGGAGCC	GCTGCCCATC	GTGTACTACG	TGGGCCGCAA	3480
GCCCAAGGTG	GAGCAGCTGT	CCAACATGAT	CGTGCGCTCC	TGCAAGTGCA	GCTGATCTAG	3540
Α .						3541

.:	(2)	INFOR	ITAM	ON F	OR S	EQ I	D NC):10:									
10		(i)	(B)	LEN TYI STI	IGTH: PE: & RANDE	ARACT 138 Amino EDNES	18 am aci SS: £	nino ld singl	acid	ls							
5. 15		(ii)	MOLE	CULI	E TYI	?E: [pepti	ide									
		(xi)	SEQU	JENCI	E DES	CRI	OIT?	V: SI	EQ II	NO:	:10:						
20		Gln 1	Leu	Ser	туг	Gly 5	Tyr	Asp	Glu	Lys	Ser 10	Thr	Gly	Gly	Ile	Ser 15	Val
25		Pro	Gly	Pro	Met 20	Gly	Pro	Ser	Gly	Pro 25	Arg	Gly	Leu	Pro	30 Gly	Pro	Pro
; 30		Gly	Ala	Pro 35	Gly	Pro	Gln	Gly	Phe 40	Gln	Gly	Pro	Pro	Gly 45	Glu	Pro	Gly
35		Glu	Pro 50	Gly	Ala	Ser	Gly	Pro 55	Met	Gly	Pro	Arg	Gly 60	Pro	Pro	Gly	Pro
40		Pro 65	Gly	Lys	Asn	Gly	Asp 70	Asp	Gly	Glu	Ala	Gly 75	Lys	Pro	Gly	Arg	Pro 80
45		Gly	Glu	Arg	Gly	Pro 85	Pro	Gly	Pro	Gln	Gly 90	Ala	Arg	Gly	Leu	Pro 95	Gly
50		Thr	Ala	Gly	Leu 100	Pro	Gly	Met	Lys	Gly 105	His	Arg	Gly	Phe	Ser	Gly	Leu

1.

ş.

	Asp	Gly	Ala	Lys	Gly	Asp	Ala	Gly	Pro	Ala	Gly	Pro	Lys	Gly	Glu	Pro
5			115					120					125			
										•						
	Gly	Ser	Pro	Gly	Glu	Asn	Gly	Ala	Pro	Gly	Gln		Gly	Pro	Arg	Gly
10		130					135					140				
			_					_		_ •	_		_			
}		Pro	Gly	Glu	Arg		Arg	Pro	GIA	Ala		GIY	Pro	Ala	Gly	
15	145					150					155					160
	R	Gly	N an) an	Glv	212	Th~	Glv	λla	פומ	Glv	Pro	Pro	G] v	Pro	Thr
	Arg	GIY	Wali	wab	165	VIG	1111	GLY	VIG	170	GLY	110		GLY	175	1111
20					203											
•	Gly	Pro	Ala	Gly	Pro	Pro	Gly	Phe	Pro	Gly	Ala	Val	Gly	Ala	Lys	Gly
				180			•		185	-				190	-	-
25																
	Glu	Ala	Gly	Pro	Gln	Gly	Pro	Arg	Gly	Ser	Glu	Gly	Pro	Gln	Gly	Val
			195					200					205			
30																
	Arg	Gly	Glu	Pro	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Ala	Ala	Gly	Pro	Ala
		210					215					220				
35											_					_
	•		Pro	Gly	Ala	_	Gly	Gln	Pro	Gly		Lys	Gly	Ala	Asn	Gly
	225			•		230					235					240
40	••-	Pro	a 1	T 1.0	21.	~1	21-	D	~ 1	Dho	Dvo	al.	*1-	7 ~~	G1	Dro
	Ala	PIO	GIY	116	245	GIÅ	Ala	PIO	GIY	250	PIO	GIY	ALG	Arg	255	PIO
•					243					230					233	
45	Ser	Glv	Pro	Gln	Gly	Pro	Gly	Gly	Pro	Pro	Gly	Pro	Lys	Gly	Asn	Ser
40		,		260			•	•	265		•		-	270		
F0	Gly	Glu	Pro	Gly	Ala	Pro	Gly	Ser	Lys	Gly	Asp	Thr	Gly	Ala	Lys	Gly
50			275					280	•				285			

i.	Glu	Pro	Gly	Pro	Val	Gly	Val	Gln	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Glu
5		290					295					300				
٠.	Glu	Gly	Lys	Arg	Gly	Ala	Arg	Gly	Glu	Pro	Gly	Pro	Thr	Gly	Leu	Pro
10	305					310					315					320
,	Gly	Pro	Pro	Gly	Glu	Arg	Gly	Gly	Pro	Gly	Ser	Arg	Gly	Phe	Pro	Gly
15					325					330					335	
	Ala	Asp	Gly	Val	Ala	Gly	Pro	Lys	Gly	Pro	Ala	Gly	Glu	Arg	Gly	Ser
				340					345					350		
20				_ •	_ _						_	_				
	Pro	Gly	Pro 355	Ala	Gly	Pro	Lys		Ser	Pro	Gly	Glu		Gly	Arg	Pro
			333					360					365			
25	Gly	Glu	Ala	Gly	Leu	Pro	Gly	Ala	Lys	Gly	Leu	Thr	Gly	Ser	Pro	Gly
Je.		370					375					380				Ī
30																
30	Ser	Pro	Gly	Pro	Asp	Gly	Lys	Thr	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Gln
	385					390					395					400
35	Asp	Gly	Arq	Pro	Glv	Pro	Pro	Glv	Pro	Pro	Glv	Ala	Arg	Glv	Gln	Ala
	•	•	J		405					410			3		415	
40	Gly	Val	Met	Gly	Phe	Pro	Gly	Pro	Lys	Gly	Ala	Ala	Gly	Glu	Pro	Gly
				420					425					430		
	Tva	בות	Glv	œ1	7~~	Gl v	1701	Dwa		D===	D==	~ 1	23-	17.3	0 1	D
45	nya	ALG	435	Glu	AL 9	Gly	Val	440	GIY	PIO	PIO	GIY	445	val	GΙΥ	PIO
													- 13			
	Ala	Gly	Lys	Asp	Gly	Glu	Ala	Gly	Ala	Gln	Gly	Pro	Pro	Gly	Pro	Ala
50		450					455					460				

	Gly	Pro	Ala	Gly	Glu	Arg	Gly	Glu	Gln	Gly	Pro	Ala	Gly	Ser	Pro	Gly
5	465					470					475					480
s .	Phe	Gln	Gly	Leu	Pro	Gly	Pro	Ala	Gly	Pro 490	Pro	Gly	Glu	Ala	Gly 495	Lys
.10																
į.	Pro	Gly	Glu	Gln 500	Gly	Val	Pro	Gly	Asp 505	Leu	Gly	Ala	Pro	Gly 510	Pro	Ser
15																
	Gly	Ala	Arg 515	Gly	Glu	Arg	Gly	Phe 520	Pro	Gly	Glu	Arg	Gly 525	Val	Gln	Gly
20			-													
	Pro	Pro 530	Gly	Pro	Ala	Gly	Pro 535	Arg	Gly	Ala	Asn	Gly 540	Ala	Pro	Gly	Asn
25	Asp	Gly	Ala	Lys	Gly	Asp	Ala	Gly	Ala	Pro	Gly	Ala	Pro	Gly	Ser	Gln
	545					550					555					560
30	Gly	Ala	Pro	Gly	Leu 565	Gln	Gly	Met	Pro	Gly 570	Glu	Arg	Gly	Ala	Ala 575	Gly
٠.													1			
35 '	Leu	Pro	Gly	Pro 580	Lys	Gly	Asp	Arg	Gly 585	Asp	Ala	Gly	Pro	Lys 590	Gly	Ala
40· ·	Asp	Gly	Ser 595	Pro	Gly	Lys	Asp	Gly 600	Val	Arg	Gly	Leu	Thr 605	Gly	Pro	Ile
45	Gļy	Pro 610		Gly	Pro	Ala	Gly 615		Pro	Gly	Asp	Lys 620		Glu	Ser	Gly
	Pro	Ser	Gly	Pro	Ala	Gly	Pro	Thr	Gly	Ala	Arg	Gly	Ala	Pro	Gly	Asp
50	625					630					635					640

55

٠,

	Arg	Gly	Glu	Pro	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Phe	Ala	Gly	Pro	Pro
5					645					650					655	
			•	a 1	~ 1		~ 3		•		01	D	01			~-
	GIA	Ala	qea	660	Gln	Pro	GIÀ	Ala	ьув 665	GIÀ	GIU	Pro	GIA	670	Ala	GIŸ
10				800					003					0,0		
	Ala	Lvs	Gly	Asp	Ala	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Pro	Ala	Gly	Pro
	••	_• -	675	•		•		680	•			-	685			
15																
	Pro	Gly	Pro	Ile	Gly	Asn	Va1	Gly	Ala	Pro	Gly	Ala	Lys	Gly	Ala	Arg
		690					695					700				
20																
	Gly	Ser	Ala	Gly	Pro	Pro	Gly	Ala	Thr	Gly	Phe	Pro	Gly	Ala	Ala	•
	705					710					715					720
25		•		_		~		_	~ 1			a 1		D	~ 1	
•	Arg	Val	GIÀ	Pro	Pro 725	GIĀ	Pro	ser	GIÀ	730	AIA	GIĀ	Pro	Pro	735	Pro
					123					730					/33	
30	Pro	Gly	Pro	Ala	Gly	Lys	Glu	Gly	Gly	Lys	Gly	Pro	Arg	Gly	Glu	Thr
,		-		740		•			745	_				750		
35	Gly	Pro	Ala	Gly	Arg	Pro	Gly	Glu	Val	Gly	Pro	Pro	Gly	Pro	Pro	Gly
			755					760					765			
40	Pro		Gly	Glu	Lys	Gly			Gly	Ala	qaA		Pro	Ala	Gly	Ala
		770					775					780				
	Pro	Gly	ጥኮሎ	Pro	Gly	Pro	Gln	Glv	Tla	Δla	Glv	Gln	Ara	Glv	Va1	Val
45	785	GIY	1111	FIU	Gry	790	G111	Gly	***	λια	795	J	 9	GLy	V4.	800
										•						
	Gly	Leu	Pro	Gly	Gln	Arg	Gly	Glu	Arg	Gly	Phe	Pro	Gly	Leu	Pro	Gly
50					805					810					815	

, 5	Pro	Ser	Gly		Pro	Gly	Lys	Gln	_	Pro	Ser	Gly	Ala		Gly	Glu
				820					825	•				830		
	Arg	Gly	Pro	Pro	Gly	Pro	Met	Gly	Pro	Pro	Gly	Leu	Ala	Gly	Pro	Pro
10			835					840					845			
	Gly	Glu	Ser	Gly	Arg	Glu	дìу	Ala	Pro	Gly	Ala	Glu	Gly	Ser	Pro	Gly
.15		850		,			855					860				
	Arg	Asp	Gly	Ser	Pro	Gly	Ala	Lys	Gly	Asp	Arg	Gly	Glu	Thr	Gly	Pro
20	865					870					875					880
	Ala	Gly	Pro	Pro	Gly	Ala	Pro	Gly	Ala	Pro	Gly	Ala	Pro	Gly	Pro	Val
25					885					890					895	
	Gly	Pro	Ala	Gly	Lys	Ser	Gly	Asp	Arg	Gly	Glu	Thr	Gly	Pro	Ala	Gly
20				900					905					910		
30	Pro	Ala	Gly	Pro	Val	Gly	Pro	Ala	Gly	Ala	Arg	Gly	Pro	Ala	Gly	Pro
			915					920					925			
35	Gln	Gly	Pro	Arg	Gly	Asp	Lys	Gly	Glu	Thr	Gly	Glu	Gln	Gly	Asp	Arg
		930	•				935					940				
40	Gly	Ile	Lys	Gly	His	Arg	Gly	Phe	Ser	Gly	Leu	Gln	Gly	Pro	Pro	Gly
	945					950					955					960
45	Pro	Pro	Gly	Ser	Pro	Gly	Glu	Gln	Gly	Pro	Ser	Gly	Ala	Ser	Gly	Pro
٠.					965					970					975	
50	Ala	Gly	Pro	Arg	Gly	Pro	Pro	Gly	Ser	Ala	Gly	Ala	Pro	Gly	Lys	Asp
		-		980	•			-	985		-			990		

	Gly	Leu	Asn	Gly	Leu	Pro	Gly	Pro	Ile	Gly	Pro	Pro	Gly	Pro	Arg	Gly
5			995					1000)				1005	•		
:	Arg	Thr	Gly	Asp	Ala	Gly	Pro	Val	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Pro
10		1010)				1015	5				1020)			
:	Pro	Gly	Pro	Pro	Gly	Pro	Pro	Ser	Ala	Gly	Phe	Asp	Phe	Ser	Phe	Leu
15	1025	5				1030)				1035	5				1040
	Pro	Gln	Pro	Pro	Gln	Glu	Lys	Ala	His	Asp	Gly	Gly	Arg	Tyr	Tyr	Arg
20					104	5				1050)				1055	5
	Ala	Arg	Ser	Asp	Glu	Ala	Ser	Gly	Ile	Gly	Pro	Glu	Val	Pro	Asp	Asp
				106	0				1065	5				1070)	
.25		3.00	Dha	C1	Dro	C0~	T 011	Clv	Dro	T/a l	Ove	Pro	Dhe	Arg	Cva	Gl n
že.	Arg	Asp	107		PIO	361	Deu	108		V41	Cys	210	1089		Cys	GIII
30			20						-							
	Cys	His	Leu	Arg	Val	Val	Gln	Сув	Ser	Asp	Leu	Gly	Leu	Asp	Lys	Val
O-		109	0				109	5				110	0			
35	Pro	Lys	Asp	Leu	Pro	Pro	Asp	Thr	Thr	Leu	Leu	Asp	Leu	Gln	Asn	Asn
n:	110	5				111	0				111	5				1120
40	Lys	Ile	Thr	Glu	Ile	Lys	Asp	Gly	Asp	Phe	Lys	Asn	Leu	Lys	Asn	Leu
					112	5				113	0				113	5
45	His	Ala	Leu	Ile	Leu	Val	Asn	Asn	Lys	Ile	Ser	Lys	Val	Ser	Pro	Gly
				114	0				114	5				115	0	
50	Ala	Phe	Thr	Pro	Leu	Val	Lys	Leu	Glu	Arg	Leu	Tyr	Leu	Ser	Lys	Asn
			115	5				116	0				116	5		
x																

	Gln	Leu	Lys	Glu	Leu	Pro	Glu	Lys	Met	Pro	Lys	Thr	Leu	Gln	Glu	Leu
5		1170	•				1175	•				1180)			
	Arg	Ala	His	Glu	Asn	Glu	Ile	Thr	Lys	Val	Arg	Lys	Val	Thr	Phe	Asn
10	1185	5				1190)				1195	5				1200
	Gly	Leu	Asn	Gln	Met	Ile	Va1	Ile	Glu	Leu	Gly	Thr	Asn	Pro	Leu	Lys
15					1205	5				1210)				1215	5
	Ser	Ser	Glv	Ile	Glu	Asn	Gly	Ala	Phe	Gln	Gly	Met	Lys	Lys	Leu	Ser
20			•	1220			•		1225		-		-	1230		
	 .	73.0	N	71.	71 -	N an	Mhw	N an	Tla	Th-	Car	710	Dro	Gl n	alv	Lou
	тут	He	1239		AIA	Asp	THE	1240		ini	261	116	1245		GIY	Leu
25																
	Pro	Pro	Ser	Leu	Thr	Glu			Leu	Asp	Gly			Ile	Ser	Arg
		1250)				125	5				1260	0			
30	Val	Asp	Ala	Ala	Ser	Leu	Lys	Gly	Leu	Asn	Asn	Leu	Ala	Lys	Leu	Gly
,	126	_				127	_	•			127					1280
35																
	Leu	Ser	Phe	Asn	Ser		Ser	Ala	Val	Asp 129		Gly	Ser	Leu	Ala 129	Asn 5
						_										
40 .	Thr	Pro	His	Leu	Arg	Glu	Leu	His	Leu	Ąsp	Asn	Asn	Lys	Leu	Thr	Arg
				130	0				130	5				131	0	
45	Val	Pro	Gly	Gly	Leu	Ala	Glu	His	Lys	Tyr	Ile	Gln	Val	Val	Tyr	Leu
			131	5				132	0				132	5		
50	His	Asn	Asn	Asn	Ile	Ser	Val	Val	Glv	Ser	Ser	asa	Phe	Cvs	Pro	Pro
•		133					133		1			134		- ,		

5	Gly 1345		Asn	Thr	Lys	Lys 1350		Ser	Tyr	Ser	Gly 1355		Ser	Leu	Phe	Ser 1360
10	Asn	Pro '	Val		Tyr 1365		Glu	Ile	Gln	Pro 1370		Thr	Phe	Arg	Cys 1375	
15	Tyr	Val		Ser 1380		Ile	Gln	Leu	Gly 1385		Tyr	Lys				
(2)	INFOR	ITAM	ON F	OR S	EQ 1	D NO):11:								,	
20	(i)	(A)	LEN	GTH :	110		STICS nino		ls							
25		(C)	STR	ANDE	EDNES		ingl	.e		-						
30	(ii)	MOLE	CULE	TYI	PE:]	pept:	ide									
	(xi)	SEQU	ÆNCE	E DES	CRI	PTIO	N: SI	EQ II	ои c	:11:						
35	Gln 1	Leu	Ser	Tyr	Gly 5	Tyr	Asp	Glu	Lys	Ser	Thr	Gly	Gly	Ile	Ser 15	Val
40	Pro	Gly	Pro	Met 20	Gly	Pro	Ser	Gly	Pro 25	Arg	Gly	Leu	Pro	Gly 30	Pro	Pro
45	Gļy	Ala	Pro 35	Gly	Pro	Gln	Gly	Phe 40	Gln	Gly	Pro	Pro	Gly 45	Glu	Pro	Gly
50	Glu	Pro 50	Gly	Ala	Ser	Gly	Pro 55	Met	Gly	Pro	Arg	Gly 60	Pro	Pro	Gly	Pro

.	Pro	Gly	Lys	Asn	Gly	Asp	Asp	Gly	Glu	Ala	Gly	Lys	Pro	Gly	Arg	Pro
5	65					70					75					80
5	Gly	Glu	Arg	Gly	Pro	Pro	Gly	Pro	Gln	Gly	Ala	Arg	Gly	Leu	Pro	Gly
10					85					90					95	
	Thr	Ala	Gly	Leu	Pro	Gly	Met	Lys	Gly	His	Arg	Gly	Phe	Ser	Gly	Leu
15				100			,		105					110		
	qaA	Gly	Ala	Lys	Gly	Asp	Ala	Gly	Pro	Ala	Gly	Pro	Lys	Gly	Glu	Pro
20			115					120					125			
	Gly	Ser	Pro	Gly	Glu	Asn	Gly	Ala	Pro	Gly	Gln	Met	Gly	Pro	Arg	Gly
		130					135					140		,		
25						_			_	_					_	
		Pro	Gly	Glu	Arg	-	Arg	Pro	Gly	Ala		Gly	Pro	Ala	Gly	Ala
ş	145					150					155					160
30	Ara	Glv	Asn	Asp	Glv	Ala	Thr	Glv	Ala	Ala	Glv	Pro	Pro	Glv	Pro	Thr
	1129	U _j			165		****			170	J-1			1	175	
8 -																
35	Gly	Pro	Ala	Gly	Pro	Pro	Gly	Phe	Pro	Gly	Ala	Val	Gly	Ala	Lys	Gly
				180					185					190		
40	Glu	Ala	-	Pro	Gln	Gly	Pro	Arg	Gly	Ser	Glu	Gly	Pro	Gln	Gly	Val
			195					200					205			
45	Arg	Gly	Glu	Pro	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Ala	Ala	Gly	Pro	Ala
		210					215					220				
50	Gly	Asn	Pro	Gly	Ala	qaA	Gly	Gln	Pro	Gly	Ala	Lys	Gly	Ala	Asn	Gly
	225					230					235					240

	Ala	Pro	Gly	Ile	Ala	Gly	Ala	Pro	Gly	Phe	Pro	Gly	Ala	Arg	Gly	Pro
5					245					250					255	
3										•						
	Ser	Gly	Pro	Gln	Gly	Pro	Gly	Gly	Pro	Pro	Gly	Pro	Lys	Gly	Asn	Ser
· ·				260					265					270		
10																
	Gly	Glu	Pro	Gly	Ala	Pro	Gly	Ser	Lys	Gly	Asp	Thr	Gly	Ala	Lys	Gly
•			275					280					285			
15																
	Glu	Pro	Gly	Pro	Val	Gly	Val	Gln	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Glu
۲.		290					295					300				
20																
	Glu	Gly	Lys	Arg	Gly	Ala	Arg	Gly	Glu	Pro	Gly	Pro	Thr	Gly	Leu	Pro
	305					310					315					320
25																
25	Gly	Pro	Pro	Gly	Glu	Arg	Gly	Gly	Pro	Gly	Ser	Arg	Gly	Phe	Pro	Gly
					325					330					335	
30	Ala	Asp	Gly	Val	Ala	Gly	Pro	Lys	Gly	Pro	Ala	Gly	Glu	Arg	Gly	Ser
				340					345					350		
D (
35	Pro	Gly	Pro	Ala	Gly	Pro	Lys	Gly	Ser	Pro	Gly	Glu	Ala	Gly	Arg	Pro
			355					360					365			
40°	Gly	Glu	Ala	Gly	Leu	Pro	Gly	Ala	Lys	Gly	Leu	Thr	Gly	Ser	Pro	Gly
· ·		370					375					380				
45	Ser	Pro	Gly	Pro	Asp	Gly	Lys	Thr	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Gln
40	385					390					395					400
	Asp	Gly	Arg	Pro	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Ala	Arg	Gly	Gln	Ala
50					405					410					415	

	Gly	Val	Met	Gly	Phe	Pro	Gly	Pro	Lys	Gly	Ala	Ala	Gly	Glu	Pro	Gly
5				420					425					430		
	Lys	Ala	Gly	Glu	Arg	Gly	Val	Pro	Gly	Pro	Pro	Gly	Ala	Val	Gly	Pro
10			435					440					445			
	Ala	Gly	Lys	Asp	Gly	Glu	Ala	Gly	Ala	Gln	Gly	Pro	Pro	Gly	Pro	Ala
15		450					455					460				
	Gly	Pro	Ala	Gly	Glu	_	Gly	Glu	Gln	Gly		Ala	Gly	Ser	Pro	-
20	465					470					475					480
	Phe	Gln	Gly	Leu	Pro	Gly	Pro	Ala	Gly	Pro	Pro	Gly	Glu	Ala	Gly	Lys
25					485					490					495	
	Pro	Gly	Glu	Gln	Gly	Val	Pro	Gly	Asp	Leu	Gly	Ala	Pro	Gly	Pro	Ser
.:				500					505					510		
30	Gly	Ala	Arg	Gly	Glu	Arg	Gly	Phe	Pro	Gly	Glu	Arg	Gly	Val	Gln	Gly
			515					520					525			
35	Pro	Pro	Gly	Pro	Ala	Gly	Pro	Arg	Gly	Ala	Asn	Gly	Ala	Pro	Gly	Asn
		530					535					540				
40	•	63		Lys	~1	1	N1-	63	31 -	D=0	C111	77.0	Pro	alv	Sar	Gl n
	545	GIY	ALA	пåв	GIÅ	550	Ala	GIÅ	ALA	PIO	555	ALG	PIO	Gly	Ser	560
45	Gly	Ala	Pro	Gly	Leu 565	Gln	Gly	Met	Pro	Gly 570	Glu	Arg	Gly	Ala	Ala 575	Gly
	T 0	Dwo	<i>a</i> 1••	Pro	Tara	GI.) Acr		Gl··	۸an	- ומ	G] v	Pro	T.ve	alv	alα
50	Dea	£10	GLY	580		OLY	vaħ	ura	585		n.a	~~ <i>y</i>		590		

	Asp	Gly	Ser	Pro	Gly	Lys	qaA	Gly	Val	Arg	Gly	Leu	Thr	Gly	Pro	Ile
5			595					600					605			
			_		_					•			_	_		
10	Gly		Pro	Gly	Pro	Ala		Ala	Pro	Gly	Asp		Gly	Glu	Ser	Gly
10		610					615					620				
	Pro	Ser	Glv	Pro	Ala	Glv	Pro	Thr	Glv	Δla	Ara	Glv	Δla	Pro	G) v	3 an
15	625	J C2	O.J			630		••••	OL,	*****	635	0.1	nzu	110	GIY	640
13																010
	Arg	Gly	Glu	Pro	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Phe	Ala	Gly	Pro	Pro
20					645					650					655	
20																
	Gly	Ala	Asp	Gly	Gln	Pro	Gly	Ala	Lys	Gly	Glu	Pro	Gly	qaA	Ala	Gly
25				660					665		٠,			670		
	_ •	_		_			_	_		_			_			
	Ala	Lys	675	Asp	Ala	GΙΆ	Pro		GIA	Pro	Ala	Gly		Ala	Gly	Pro
30			0/3					680					685			
	Pro	Gly	Pro	Ile	Gly	Asn	Val	Glv	Ala	Pro	Glv	Ala	Lvs	Gly	Ala	Ara
·		690			•		695	,			2	700	-1-	1		•5
35 [†]																
	Gly	Ser	Ala	Gly	Pro	Pro	Gly	Ala	Thr	Gly	Phe	Pro	Gly	Ala	Ala	Gly
74	705					710					715					720
40																
	Arg	Val	Gly	Pro		Gly	Pro	Ser	Gly	Asn	Ala	Gly	Pro	Pro	Gly	Pro
					725					730					735	
45	Pro	G1v	Pro	212	GT v	Tare	cl.,	G1	~1	T 1 1 4	~1	Dwa	3	Gly	63	ml
	PIO	GIY	PIO	740	GIY	пув	GIU	GIY	745	гув	GIY	PIO	Arg	750	GIU	Tnr
									, 43					750		
50	Gly	Pro	Ala	Gly	Arg	Pro	Gly	Glu	Val	Gly	Pro	Pro	Gly	Pro	Pro	Gly
			755					760		-			765			-
;																

	Pro	Ala	Gly	Glu	Lys	Gly	Ser	Pro	Gly	Ala	Asp	Gly	Pro	Ala	Gly	Ala
5		770					775					780				
٠,																
	Pro	Gly	Thr	Pro	Gly	Pro	Gln	Gly	Ile	Ala	Gly	Gln	Arg	Gly	Val	Val
10	785					790					795					800
	Gly	Leu	Pro	Gly	Gln	Arg	Gly	Glu	Arg	Gly	Phe	Pro	Gly	Leu	Pro	Gly
15					805					810					815	
										_	_			_		
•	Pro	Ser	Gly		Pro	Gly	Lys	Gln	_	Pro	Ser	Glγ	Ala		Gly	Glu
20				820					825					830		
	•	a 1	D	D	al	D	Wa b	a 1	D	Dwa	a1	T 011	710	~1	Pro	D
	Arg	GIY	835	PIO	GIY	PIO	Mec	840	PIO	PIO	GIY	neu	845	GIY	PIO	PLO
25			633					040					013			
	G] v	Glu	Ser	Glv	Ara	Glu	Glv	Ala	Pro	Glv	Ala	Glu	Glv	Ser	Pro	Glv
1	0. y	850	-	,	3		855			2		860				2
30																
	Arg	Asp	Gly	Ser	Pro	Gly	Ala	Lys	Gly	Asp	Arg	Gly	Glu	Thr	Gly	Pro
•	865					870					875					880
35																
30	Ala	Gly	Pro	Pro	Gly	Ala	Pro	Gly	Ala	Pro	Gly	Ala	Pro	Gly	Pro	Val
					885					890					895	
40 ⁻	Gly	Pro	Ala	Gly	Lys	Ser	Gly	Asp	Arg	Gly	Glu	Thr	Gly	Pro	Ala	Gly
				900					905					910		
45	Pro	Ala	_		Val	Gly	Pro		Gly	Ala	Arg	Gly		Ala	Gly	Pro
			915					920					925			
			_	•	~.	_	_	a :	a 3	~ 1	a 3 -	43 :	~ 3	63 -	•	•
50 ⁻	Gln	•		Arg	GIÀ	Asp	_	GΙΆ	Glu	Tnr	GIÀ		GIn	GTÅ	Ąsp	Arg
		930					935					940				

55

•:

	Gly	Ile	Lys	Gly	His	Arg	Gly	Phe	Ser	Gly	Leu	Gln	Gly	Pro	Pro	Gly
5	945					950					955					960
													_			
	Pro	Pro	Gly	Ser		Gly	Glu	Gln	Gly		Ser	Gly	Ala	Ser	Gly	Pro
0					965					970					975	
	••-	01	77-0	7.50	Gly	Dro	Dro	Gly.	Car	בומ	Glv	λla	Pro	Glv	Lys	7 co
_	Ala	GIY	PLO	980	GIŞ	PIO	PLO	GIY	985	ALG	Q1y	714		990	nys	veb
5				500					,,,					,,,		
	Gly	Leu	Asn	Gly	Leu	Pro	Gly	Pro	Ile	Gly	Pro	Pro	Gly	Pro	Arg	Gly
	-		995					1000)				100	5		
20																
	Arg	Thr	Gly	Asp	Ala	Gly	Pro	Val	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Pro
		1010	0				1015	5				102	0			
?5														•		
	Pro	Gly	Pro	Pro	Gly	Pro	Pro	Ser	Ala	Gly	Phe	Asp	Phe	Ser	Phe	Leu
	1025	5				103	0				103	5				1040
30																
	Pro	Gln	Pro	Pro			Lys	Ala	His			Gly	Arg	Tyr	Tyr	
					104	5				105	0				105	5
35		_	•		•	S	•	D	D	3	mh sa	mh w	T ou	7 011	A an	T ou
	Ala	Arg	ser			Asp	Leu	Pro	106		THE	THE	Leu	107	Asp o	Den
				106	U				100	.				107	•	
10	Gln	Agn	Asn	Lvs	Ile	Thr	Glu	Ile	Lvs	Asp	Glv	Asp	Phe	Lvs	Asn	Leu
			107	-	-,			108				•	108			
15·	Lys	Asn	Leu	His	Ala	Leu	Ile	Leu	Val	Asn	Asn	Lys	Ile	Ser	Lys	Val
	_	109	0			•	109	5				110	0			
•																
50	Ser	Pro	Gly	•												
	110	5														
£.																

55·

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 4167 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

20

25

30

35

40

45

50

5

10

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CAGCTGTCTT ATGGCTATGA TGAGAAATCA ACCGGAGGAA TTTCCGTGCC TGGCCCCATG 60 GGTCCCTCTG GTCCTCGTGG TCTCCCTGGC CCCCTGGTG CACCTGGTCC CCAAGGCTTC 120 CAAGGTCCCC CTGGTGAGCC TGGCGAGCCT GGAGCTTCAG GTCCCATGGG TCCCCGAGGT 180 CCCCCAGGTC CCCCTGGAAA GAATGGAGAT GATGGGGAAG CTGGAAAACC TGGTCGTCCT 240 GGTGAGCGTG GGCCTCCTGC GCCTCAGGGT GCTCGAGGAT TGCCCGGAAC AGCTGGCCTC 300 CCTGGAATGA AGGGACACAG AGGTTTCAGT GGTTTGGATG GTGCCAAGGG AGATGCTGGT 360 CCTGCTGGTC CTAAGGGTGA GCCTGGCAGC CCTGGTGAAA ATGGAGCTCC TGGTCAGATG 420 GGCCCCCGTG GCCTGCCTGG TGAGAGAGGT CGCCCTGGAG CCCCTGGCCC TGCTGGTGCT 480 CGTGGAAATG ATGGTGCTAC TGGTGCTGCC GGGCCCCCTG GTCCCACCGG CCCCGCTGGT 540 CCTCCTGGCT TCCCTGGTGC TGTTGGTGCT AAGGGTGAAG CTGGTCCCCA AGGGCCCCGA 600

. .

GGCTCTGAAG	GTCCCCAGGG	TGTGCGTGGT	GAGCCTGGCC	CCCCTGGCCC	TGCTGGTGCT	660
GCTGGCCCTG	CTGGAAACCC	TGGTGCTGAT	GGACAGCCȚG	GTGCTAAAGG	TGCCAATGGT	720
GCTCCTGGTA	TTGCTGGTGC	TCCTGGCTTC	CCTGGTGCCC	GAGGCCCCTC	TGGACCCCAG	780
GGCCCCGGCG	GCCCTCCTGG	TCCCAAGGGT	AACAGCGGTG	AACCTGGTGC	TCCTGGCAGC	840
AAAGGAGACA	CTGGTGCTAA	GGGAGAGCCT	GGCCCTGTTG	GTGTTCAAGG	ACCCCCTGGC	900
CCTGCTGGAG	AGCAAGGAAA	GCGAGGAGCT	CGAGGTGAAC	CCGGACCCAC	TGGCCTGCCC	960
GGACCCCCTG	GCGAGCGTGG	TGGACCTGGT	AGCCGTGGTT	TCCCTGGCGC	AGATGGTGTT	1020
GCTGGTCCCA	AGGGTCCCGC	TGGTGAACGT	GGTTCTCCTG	GCCCCGCTGG	CCCCAAAGGA	1080
TCTCCTCGTG	AAGCTGGTCG	TCCCGGTGAA	GCTGGTCTGC	CTGGTGCCAA	GGGTCTGACT	1140
GGAAGCCCTG	GCAGCCCTGG	TCCTGATGGC	AAAACTGGCC	CCCCTGGTCC	CGCCGGTCAA	1200
GATGGTCGCC	CCGGACCCCC	AGGCCCACCT	GGTGCCCGTG	GTCAGGCTGG	TGTGATGGGA	1260
TTCCCTGGAC	CTAAAGGTGC	TGCTCGAGAG	CCCGGCAAGG	CTGGAGAGCG	AGGTGTTCCC	1320
GGACCCCCTC	GCGCTGTCGG	TCCTGCTGGC	AAAGATGGAG	AGGCTGGAGC	TCAGGGACCC	1380
CCTGGCCCTG	CTGGTCCCGC	TGGCGAGAGA	GGTGAACAAG	GCCCTGCTGG	CTCCCCGGA	1440
TTCCAGGGTC	TCCCTGGTCC	TGCTGGTCCT	CCAGGTGAAG	CAGGCAAACC	TGGTGAACAG	1500
GGTGTTCCTG	GAGACCTTGG	CGCCCCTGGC	CCCTCTGGAG	CAAGAGGCGA	GAGAGGTTTC	1560
CCTGGCGAGC	GTGGTGTGCA	AGGTCCCCCT	GGTCCTGCTG	GACCCCGAGG	GGCCAACGGT	1620

`20

	GCTCCCGCCA	ACGATGCTGC	TAAGGGTGAT	GCTGGTGCCC	CTGGAGCTCC	CGGTAGCCAG	1680
	GGCGCCCTG	GCCTTCAGGG	AATGCCTGGT	GAACGTGGŢG	CAGCTGGTCT	TCCAGGGCCT	1740
	AAGGGTGACA	GAGGTGATGC	TGGTCCCAAA	GGTGCTGATG	GCTCTCCTGG	CAAAGATGGC	1800
	GTCCGTGGTC	TGACCGACCC	CATTGGTCCT	CCTGGCCCTG	CTGGTGCCCC	TGGTGACAAG	1860
	GGTGAAAGTG	GTCCCAGCGG	CCCTGCTGGT	CCCACTGGAG	CTCGTGGTGC	CCCCGGAGAC	1920
-	CGTGGTGAGC	CTGGTCCCCC	CGGCCCTGCT	GGCTTTGCTG	GCCCCCTGG	TGCTGACGGC	1980
	CAACCTGGTG	CTAAAGGCGA	ACCTGGTGAT	GCTGGTGCCA	AAGGCGATGC	TGGTCCCCCT	2040
	GGGCCTGCCG	GACCCGCTGG	ACCCCCTGGC	CCCATTGGTA	ATGTTGGTGC	TCCTGGAGCC	2100
	AAACGTGCTC	GCGGCAGCGC	TGGTCCCCCT	GGTGCTACTG	GTTTCCCTGG	TGCTGCTGGC	2160
	CGAGTCGGTC	CTCCTGGCCC	CTCTGGAAAT	GCTGGACCCC	CTGGCCCTCC	TGGTCCTGCT	2220
	GGCAAAGAAG	GCGGCAAAGG	TCCCCGTGGT	GAGACTGGCC	CTGCTGGACG	TCCTGGTGAA	2280
	GTTGGTCCCC	CTGGTCCCCC	TGGCCCTGCT	GGCGAGAAAG	GATCCCCTGG	TGCTGATGGT	2340
	CCTGCTGGTG	CTCCTGGTAC	TCCCGGGCCT	CAAGGTATTG	CTGGACAGCG	TGGTGTGGTC	2400
	GGCCTGCCTG	GTCAGAGAGG	AGAGAGAGGC	TTCCCTGGTC	TTCTTGGCCC	CTCTGGTGAA	2460
	CCTGGCAAAC	AAGGTCCCTC	TGGAGCAAGT	GGTGAACGTG	GTCCCCCCGG	TCCCATGGGC	2520
	CCCCTGGAT	TGGCTGGACC	CCCTGGTGAA	TCTGGACGTG	AGGGGGCTCC	TGCTGCCGAA	2580
	GGTTCCCCTG	GACGAGACGG	TTCTCCTGGC	GCCAAGGGTG	ACCGTGGTGA	GACCGGCCCC	2640

- 5

:10

55

<u>్స్</u>

÷5	GCTGGACCCC	CTGGTGCTCC	TGGTGCTCCT	GGTGCCCCTG	GCCCCGTTGG	CCCTGCTGGC	2700
	AAGAGTGGTG	ATCGTGGTGA	GACTGGTCCT	GCTGGTCCÇG	CCGGTCCCGT	CGGCCCCGCT	2760
:10	GGCGCCCGTG	GCCCCGCCGG	ACCCCAAGGC	CCCCGTGGTG	ACAAGGGTGA	GACAGGCGAA	2820
	CAGGGCGACA	GAGGCATAAA	GGGTCACCGT	GGCTTCTCTG	GCCTCCAGGG	TCCCCCTGGC	2880
15	CCTCCTGGCT	CTCCTGGTGA	ACAAGGTCCC	TCTGGAGCCT	CTGGTCCTGC	TGGTCCCCGA	2940
20	GGTCCCCCTG	GCTCTGCTGG	TGCTCCTGGC	AAAGATGGAC	TCAACGGTCT	CCCTGGCCCC	3000
	ATTGGGCCCC	CTGGTCCTCG	CGGTCGCACT	GGTGATGCTG	GTCCTGTTGG	TCCCCCGGC	3-060
25	CCTCCTGGAC	CTCCTGGTCC	CCCTGGTCCT	CCCAGCGCTG	GTTTCGACTT	CAGCTTCCTC	3120
	CCCCAGCCAC	CTCAAGAGAA	GGCTCACGAT	GGTGGCCGCT	ACTACCGGGC	TAGATCCGAT	3180
30	GAGGCTTCTG	GGATAGCCCC	AGAAGTTCCT	GATGACCGCG	ACTTCGAGCC	CTCCCTAGGC	3240
35	CCAGTGTGCC	CCTTCCGCTG	TCAATGCCAT	CTTCGAGTGG	TCCAGTGTTC	TGATTTGGGT	3300
	CTGGACAAAG	TGCCAAAGGA	TCTTCCCCCT	GACACAACTC	TGCTAGACCT	GCAAAACAAC	3360
40	AAAATAACCG	AAATCAAAGA	TGGAGACTTT	AAGAACCTGA	AGAACCTTCA	CGCATTGATT	3420
45	CTTGTCAACA	ATAAAATTAG	CAAAGTTAGT	CCTGGAGCAT	TTACACCTTT	GGTGAAGTTG	3480
	GAACGACTTT	ATCTGTCCAA	GAATCAGCTG	AAGGAATTGC	CAGAAAAAAT	GCCCAAAACT	3540
50	CTTCAGGAGC	TGCGTGCCCA	TGAGAATGAG	ATCACCAAAG	TGCGAAAAGT	TACTTTCAAT	3600
	GGACTGAACC	AGATGATTGT	CATAGAACTG	GGCACCAATC	CGCTGAAGAG	CTCAGGAATT	3660

	GAAAATGGGG CTTTCCAGGG AATGAAGAAG CTCTCCTACA TCCGCATTGC TGATACCAAT	3720
:	ATCACCAGCA TTCCTCAAGG TCTTCCTCCT TCCCTTACGG AATTACATCT TGATGGCAAC	3780
10	AAAATCAGCA GAGTTGATGC AGCTAGCCTG AAAGGACTGA ATAATTTGGC TAAGTTGGGA	3840
	TTGAGTTTCA ACAGCATCTC TGCTGTTGAC AATGGCTCTC TGGCCAACAC GCCTCATCTG	3900
15	AGGGAGCTTC ACTTGGACAA CAACAAGCTT ACCAGAGTAC CTGGTGGGCT GGCAGAGCAT	3960
20	AAGTACATCC AGGTTGTCTA CCTTCATAAC AACAATATCT CTGTAGTTGG ATCAAGTGAC	4020
	TTCTGCCCAC CTGGACACAA CACCAAAAAG GCTTCTTATT CGGGTGTGAG TCTTTTCAGC	4080
25 ⁻	AACCCGGTCC AGTACTGGGA GATACAGCCA TCCACCTTCA GATGTGTCTA CGTGCGCTCT	4140
	GCCATTCAAC TCGGAAACTA TAAGTAA	4167
30	(2) INFORMATION FOR SEQ ID NO:13:	
35 ⁻	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 3349 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: single	
40	(D) TOPOLOGY: linear	
Ÿ	(ii) MOLECULE TYPE: cDNA	
45 ⁻		
50	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:	
	GGGAAGGATT TCCATTTCCC AGCTGTCTTA TGGCTATGAT GAGAAATCAA CCGGAGGAAT	60
55 '		

45[.]

50·

TTCCGTGCCT	GGCCCCATGG	GTCCCTCTGG	TCCTCGTGGT	CTCCCTGGCC	CCCCTGGTGC	120
ACCTGGTCCC	CAAGGCTTCC	AAGGTCCCCC	TGGTGAGCCT	GGCGAGCCTG	GAGCTTCAGG	180
TCCCATGGGT	CCCCGAGGTC	CCCCAGGTCC	CCCTGGAAAG	AATGGAGATG	ATGGGGAAGC	240
TGGAAAACCT	GGTCGTCCTG	GTGAGCGTGG	GCCTCCTGGG	CCTCAGGGTG	CTCGAGGATT	300
GCCCGGAACA	GCTGGCCTCC	CTGGAATGAA	GGGACACAGA	GGTTTCAGTG	GTTTGGATGG	360
TGCCAAGGGA	GATGCTGGTC	CTGCTGGTCC	TAAGGGTGAG	CCTGGCAGCC	CTGGTGAAAA	420
TGGAGCTCCT	GGTCAGATGG	GCCCCGTGG	CCTGCCTGGT	GAGAGAGGTC	GCCCTGGAGC	480
CCCTGGCCCT	GCTGGTGCTC	GTGGAAATGA	TGGTGCTACT	GGTGCTGCCG	GGCCCCCTGG	540
TCCCACCGGC	CCCGCTGGTC	CTCCTGGCTT	CCCTGGTGCT	GTTGGTGCTA	AGGGTGAAGC	600
TGGTCCCCAA	GGGCCCCGAG	GCTCTGAAGG	TCCCCAGGGT	GTGCGTGGTG	AGCCTGGCCC	660
CCCTGGCCCT	GCTGGTGCTG	CTGGCCCTGC	TGGAAACCCT	GGTGCTGATG	GACAGCCTGG	720
TGCTAAAGGT	GCCAATGGTG	CTCCTGGTAT	TGCTGGTGCT	CCTGGCTTCC	CTGGTGCCCG	780
AGGCCCCTCT	GGACCCCAGG	GCCCGGCGG	CCCTCCTGGT	CCCAAGGGTA	ACAGCGGTGA	840
ACCTGGTGCT	CCTGGCAGCA	AAGGAGACAC	TGGTGCTAAG	GGAGAGCCTG	GCCCTGTTGG	900
TGTTCAAGGA	CCCCTGGCC	CTGCTGGAGA	GGAAGGAAAG	CGAGGAGCTC	GAGGTGAACC	960
CGGACCCACT	GGCCTGCCCG	GACCCCCTGG	CGAGCGTGGT	GGACCTGGTA	GCCGTGGTTT	1020
CCCTGGCGCA	GATGGTGTTG	CTGGTCCCAA	GGGTCCCGCT	GGTGAACGTG	GTTCTCCTGG	1080

ن.

•	CCCGCTGGC	CCCAAAGGAT	CTCCTGGTGA	AGCTGGTCGT	CCCGGTGAAG	CTGGTCTGCC	1140
•	rggtgccaag	GGTCTGACTG	GAAGCCCTGG	CAGCCCTGGT	CCTGATGGCA	AAACTGGCCC	1200
•	CCTGGTCCC	GCCGGTCAAG	ATGGTCGCCC	CGGACCCCCA	GGCCCACCTG	GTGCCCGTGG	1260
٠	TCAGGCTGGT	GTGATGGGAT	TCCCTGGACC	TAAAGGTGCT	GCTGGAGAGC	CCGGCAAGGC	1320
•	TGGAGAGCGA	GGTGTTCCCG	GACCCCCTGG	CGCTGTCGGT	CCTGCTGGCA	AAGATGGAGA	1380
•	GGCTGGAGCT	CAGGGACCCC	CTGGCCCTGC	TGGTCCCGCT	GGCGAGAGAG	GTGAACAAGG	1440
	CCCTGCTGGC	TCCCCGGAT	TCCAGGGTCT	CCCTGGTCCT	GCTGGTCCTC	CAGGTGAAGC	1500
	AGGCAAACCT	GGTGAACAGG	GTGTTCCTGG	AGACCTTGGC	GCCCCTGGCC	CCTCTGGAGC	1560
	AAGAGGCGAG	AGAGGTTTCC	CTGGCGAGCG	TGGTGTGCAA	GGTCCCCTG	GTCCTGCTGG	1620
	ACCCCGAGGG	GCCAACGGTG	CTCCCGGCAA	CGATGGTGCT	AAGGGTGATG	CTGGTGCCCC	1680
	TGGAGCTCCC	GGTAGCCAGG	GCGCCCTGG	CCTTCAGGGA	ATGCCTGGTG	AACGTGGTGC	1740
	AGCTGGTCTT	CCAGGGCCTA	AGGGTGACAG	AGGTGATGCT	GGTCCCAAAG	GTGCTGATGG	1800
	CTCTCCTGGC	AAAGATGGCG	TCCGTGGTCT	GACCGGCCCC	ATTGGTCCTC	CTGGCCCTGC	1860
	TGGTGCCCCT	GGTGACAAGG	GTGAAAGTGG	TCCCAGCGGC	CCTGCTGGTC	CCACTGGAGC	1920
	TCGTGGTGCC	CCCGGAGACC	GTGGTGAGCC	TGGTCCCCC	GGCCCTGCTG	GCTTTGCTGG	1980
	CCCCCTGGT	GCTGACGGCC	AACCTGGTGC	TAAAGGCGAA	CCTGGTGATG	CTGGTGCCAA	2040
	AGGCGATGCT	GGTCCCCCTG	GGCCTGCCGG	ACCCGCTGGA	CCCCTGGCC	CCATTGGTAA	2100

:.

TGTTGGTGCT	CCTGGAGCCA	AAGGTGCTCG	CGGCAGCGCT	GGTCCCCTG	GTGCTACTGG	2160
TTTCCCTGGT	GCTGCTGGCC	GAGTCGGTCC	тсствесссс	TCTGGAAATG	CTGGACCCCC	2220
TGGCCCTCCT	GGTCCTGCTG	GCAAAGAAGG	CGGCAAAGGT	CCCCGTGGTG	AGACTGGCCC	2280
TGCTGGACGT	CCTGGTGAAG	TTGGTCCCCC	TGGTCCCCCT	GGCCCTGCTG	GCGAGAAAGG	2340
ATCCCCTGGT	GCTGATGGTC	CTGCTGGTGC	TCCTGGTACT	CCCGGGCCTC	AAGGTATTGC	2400
TGGACAGCGT	GGTGTGGTCG	GCCTGCCTGG	TCAGAGAGGA	GAGAGAGGCT	TCCCTGGTCT	2460
TCCTGGCCCC	TCTGGTGAAC	CTGGCAAACA	AGGTCCCTCT	GGAGCAAGTG	GTGAACGTGG	2520
TCCCCCCGGT	CCCATGGGCC	CCCCTGGATT	GGCTGGACCC	CCTGGTGAAT	CTGGACGTGA	2580
GGGGGCTCCT	GCTGCCGAAG	GTTCCCCTGG	ACGAGACGGT	TCTCCTGGCG	CCAAGGGTGA	2640
CCGTGGTGAG	ACCGGCCCCG	CTGGACCCCC	TGGTGCTCCT	GGTGCTCCTG	GTGCCCCTGG	2700
CCCCGTTGGC	CCTGCTGGCA	AGAGTGGTGA	TCGTGGTGAG	ACTGGTCCTG	CTGGTCCCGC	2760
CGGTCCCGTC	GGCCCCGCTG	GCGCCCGTGG	CCCCGCCGGA	CCCCAAGGCC	CCCGTGGTGA	2820
CAAGGGTGAG	ACAGGCGAAC	AGGGCGACAG	AGGCATAAAG	GGTCACCGTG	GCTTCTCTGG	2880
CCTCCAGGGT	CCCCCTGGCC	CTCCTGGCTC	TCCTGGTGAA	CAAGGTCCCT	CTGGAGCCTC	2940
TGGTCCTGCT	GGTCCCCGAG	GTCCCCCTGG	CTCTGCTGGT	GCTCCTGGCA	AAGATGGACT	3000
CAACGGTCTC	CCTGGCCCCA	TTGGGCCCCC	TGGTCCTCGC	GGTCGCACTG	GTGATGCTGG	3060
TCCTGTTGGT	cccccccccc	CTCCTGGACC	TCCTGGTCCC	CCTGGTCCTC	CCAGCGCTGG	3120

	TTTCGACTTC AGCTTCCTCC CCCAGCCACC TCAAGAGAAG GCTCACGATG GTGGCCGCTA	3180
5	CTACCGGGCT AGATCTCCAA AGGATCTTCC CCCTGACACA ACTCTGCTAG ACCTGCAAAA	3240
10	CAACAAAATA ACCGAAATCA AAGATGGAGA CTTTAAGAAC CTGAAGAACC TTCACGCATT	3300
	GATTCTTGTC AACAATAAAA TTAGCAAAGT TAGTCCTGGA TAACTGCAG	3349
15	(2) INFORMATION FOR SEQ ID NO:14:	
	(i) SEQUENCE CHARACTERISTICS:	
20	(A) LENGTH: 57 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
25	(ii) MOLECULE TYPE: cDNA	
30		
35	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:	
	ATCGAGGGAA GGATTTCAGA ATTCGGATCC TCTAGAGTCG ACCTGCAGGC AAGCTTG	57
40	(2) INFORMATION FOR SEQ ID NO:15:	
	(i) SEQUENCE CHARACTERISTICS:	
45 [°]	(A) LENGTH: 3171 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: single	
e'o	(D) TOPOLOGY: linear	
50 		
	(ii) MOLECULE TYPE: cDNA	

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

5

. 55

	CAGCTGTCTT	ATGGCTATGA	TGAGAAATCA	ACCGGAGGAA	TTTCCGTGCC	TGGCCCCATG	60
10	GGTCCCTCTG	GTCCTCGTGG	TCTCCCTGGC	CCCCCTGGTG	CACCTGGTCC	CCAAGGCTTC	120
•	CAAGGTCCCC	CTGGTGAGCC	TGGCGAGCCT	GGAGCTTCAG	GTCCCATGGG	TCCCCGAGGT	180
15	CCCCCAGGTC	CCCCTGGAAA	GAATGGAGAT	GATGGGGAAG	CTGGAAAACC	TGGTCGTCCT	240
20	GGTGAGCGTG	GGCCTCCTGG	GCCTCAGGGT	GCTCGAGGAT	TGCCCGGAAC	AGCTGGCCTC	300
	CCTGGAATGA	AGGGACACAG	AGGTTTCAGT	GGTTTGGATG	GTGCCAAGGG	AGATGCTGGT	360
25	CCTGCTGGTC	CTAAGGGTGA	GCCTGGCAGC	CCTGGTGAAA	ATGGAGCTCC	TGGTCAGATG	420
	GGCCCCCGTG	GCCTGCCTGG	TGAGAGAGGT	CGCCCTGGAG	CCCCTGGCCC	TGCTGGTGCT	480
30	CGTGGAAATG	ATGGTGCTAC	TGGTGCTGCC	GGGCCCCCTG	GTCCCACCGG	CCCCGCTGGT	540
: 35	CCTCCTGGCT	TCCCTGGTGC	TGTTGGTGCT	AAGGGTGAAG	CTGGTCCCCA	AGGGCCCCGA	600
	GGCTCTGAAG	GTCCCCAGGG	TGTGCGTGGT	GAGCCTGGCC	CCCCTGGCCC	TGCTGGTGCT	660
40 .	GCTGGCCCTG	CTGGAAACCC	TGGTGCTGAT	GGACAGCCTG	GTGCTAAAGG	TGCCAATGGT	720
	GCTCCTGGTA	TTGCTGGTGC	TCCTGGCTTC	CCTGGTGCCC	GAGGCCCCTC	TGGACCCCAG	780
45	GGCCCCGGCG	GCCCTCCTGG	TCCCAAGGGT	AACAGCGGTG	AACCTGGTGC	TCCTGGCAGC	840
 50	AAAGGAGACA	CTGGTGCTAA	GGGAGAGCCT	GGCCCTGTTG	GTGTTCAAGG	ACCCCCTGGC	900

CCTGCTGGAG	AGGAAGGAAA	GCGAGGAGCT	CGAGGTGAAC	CCGGACCCAC	TGGCCTGCCC	960
GGACCCCCTG	GCGAGCGTGG	TGGACCTGGT	AGCCGTGGTT	TCCCTGGCGC	AGATGGTGTT	1020
GCTGGTCCCA	AGGGTCCCGC	TGGTGAACGT	GGTTCTCCTG	GCCCCGCTGG	CCCCAAAGGA	1080
TCTCCTGGTG	AAGCTGGTCG	TCCCGGTGAA	GCTGGTCTGC	CTGGTGCCAA	GGGTCTGACT	1140
GGAAGCCCTG	GCAGCCCTGG	TCCTGATGGC	AAAACTGGCC	CCCCTGGTCC	CGCCGGTCAA	1200
GATGGTCGCC	CCGGACCCCC	AGGCCCACCT	GGTGCCCGTG	GTCAGGCTGG	TGTGATGGGA	1260
TTCCCTGGAC	CTAAAGGTGC	TGCTGGAGAG	CCCGGCAAGG	CTGGAGAGCG	AGGTGTTCCC	1320
GGACCCCCTG	GCGCTGTCGG	TCCTGCTGGC	AAAGATGGAG	AGGCTGGAGC	TCAGGGACCC	1380
CCTGGCCCTG	CTGGTCCCGC	TGGCGAGAGA	GGTGAACAAG	GCCCTGCTGG	CTCCCCCGGA	1440
TTCCAGGGTC	TCCCTGGTCC	TGCTGGTCCT	CCAGGTGAAG	CAGGCAAACC	TGGTGAACAG	1500
GGTGTTCCTG	GAGACCTTGG	CGCCCCTGGC	CCCTCTGGAG	CAAGAGGCGA	GAGAGGTTTC	1560
CCTGGCGAGC	GTGGTGTGCA	AGGTCCCCCT	GGTCCTGCTG	GACCCCGAGG	GGCCAACGGT	1620
GCTCCCGGCA	ACGATGGTGC	TAAGGGTGAT	GCTGGTGCCC	CTGGAGCTCC	CGGTAGCCAG	1680
GGCGCCCCTG	GCCTTCAGGG	AATGCCTGGT	GAACGTGGTG	CAGCTGGTCT	TCCAGGGCCT	1740
AAGGGTGACA	GĄGGTGATGC	TGGTCCCAAA	GGTGCTGATG	GCTCTCCTGG	CAAAGATGGC	1800
GTCCGTGGTC	TGACCGGCCC	CATTGGTCCT	CCTGGCCCTG	CTGGTGCCCC	TGGTGACAAG	1860
GGTGAAAGTG	GTCCCAGCGG	CCCTGCTGGT	CCCACTGGAG	CTCGTGGTGC	CCCCGGAGAC	1920

CGTGGTGAGC	CTGGTCCCCC	CGGCCCTGCT	GGCTTTGCTG	GCCCCCTGG	TGCTGACGGC	1980
CAACCTGGTG	CTAAAGGCGA	ACCTGGTGAT	GCTGGTGCÇA	AAGGCGATGC	TGGTCCCCCT	2040
GGGCCTGCCG	GACCCGCTGG	ACCCCCTGGC	CCCATTGGTA	ATGTTGGTGC	TCCTGGAGCC	2100
AAAGGTGCTC	GCGGCAGCGC	TGGTCCCCCT	GGTGCTACTG	GTTTCCCTGG	TGCTGCTGGC	2160
CGAGTCGGTC	CTCCTGGCCC	CTCTGGAAAT	GCTGGACCCC	CTGGCCCTCC	TGGTCCTGCT	2220
GGCAAAGAAG	GCGGCAAAGG	TCCCCGTGGT	GAGACTGGCC	CTGCTGGACG	TCCTGGTGAA	2280
GTTGGTCCCC	CTGGTCCCCC	TGGCCCTGCT	GGCGAGAAAG	GATCCCCTGG	TGCTGATGGT	2340
CCTGCTGGTG	CTCCTGGTAC	TCCCGGGCCT	CAAGGTATTG	CTGGACAGCG	TGGTGTGGTC	2400
GGCCTGCCTG	GTCAGAGAGG	AGAGAGAGGC	TTCCCTGGTC	TTCCTGGCCC	CTCTGGTGAA	2460
CCTGGCAAAC	AAGGTCCCTC	TGGAGCAAGT	GGTGAACGTG	GTCCCCCGG	TCCCATGGGC	2520
CCCCTGGAT	TGGCTGGACC	CCCTGGTGAA	TCTGGACGTG	AGGGGGCTCC	TGCTGCCGAA	2580
GGTTCCCCTG	GACGAGACGG	TTCTCCTGGC	GCCAAGGGTG	ACCGTGGTGA	GACCGGCCCC	2640
GCTGGACCCC	CTGGTGCTCC	TGGTGCTCCT	GGTGCCCCTG	GCCCCGTTGG	CCCTGCTGGC	2700
AAGAGTGGTG	ATCGTGGTGA	GACTGGTCCT	GCTGGTCCCG	CCGGTCCCGT	CGGCCCCGCT	2760
GGCGCCCGTG	GCCCCGCCGG	ACCCCAAGGC	CCCCGTGGTG	ACAAGGGTGA	GACAGGCGAA	2820
CAGGGCGACA	GAGGCATAAA	GGGTCACCGT	GGCTTCTCTG	GCCTCCAGGG	TCCCCCTGGC	2880
CCTCCTGGCT	CTCCTGGTGA	ACAAGGTCCC	TCTGGAGCCT	CTGGTCCTGC	TGGTCCCCGA	2940

	GGTCCCCCTG GCTCTGCTGG TGCTCCTGGC AAAGATGGAC TCAACGGTCT CCCTGGCCCC	3000
5	ATTGGGCCCC CTGGTCCTCG CGGTCGCACT GGTGATGCTG GTCCTGTTGG TCCCCCCGGC	3060
5- 10	CCTCCTGGAC CTCCTGGTCC CCCTGGTCCT CCCAGCGCTG GTTTCGACTT CAGCTTCCTC	3120
	CCCCAGCCAC CTCAAGAGAA GGCTCACGAT GGTGGCCGCT ACTACCGGGC T	3171
15	(2) INFORMATION FOR SEQ ID NO:16:	
	(i) SEQUENCE CHARACTERISTICS:	
20	(A) LENGTH: 1057 amino acids	
	(B) TYPE: amino acid	•
,	(C) STRANDEDNESS: single	
25	(D) TOPOLOGY: unknown	
	(ii) MOLECULE TYPE: peptide	
30		
;		
35'	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:	
	Gln Leu Ser Tyr Gly Tyr Asp Glu Lys Ser Thr Gly Gly Ile Ser Val	
	1 5 10 15	
40		
	Pro Gly Pro Met Gly Pro Ser Gly Pro Arg Gly Leu Pro Gly Pro Pro	
	20 25 30	
45		
	Gly Ala Pro Gly Pro Gln Gly Phe Gln Gly Pro Pro Gly Glu Pro Gly	
	35 40 45	
50		
	Glu Pro Gly Ala Ser Gly Pro Met Gly Pro Arg Gly Pro Pro Gly Pro	
	50 55 60	
55		

 $\vec{\beta}$

5		Gly	Lys	Asn	Gly		Asp	Gly	Glu	Ala		Гåа	Pro	Gly	Arg	
	65					70					75					80
	Glv	Glu	Arg	Glv	Pro	Pro	Gly	Pro	Gln	Gly	Ala	Arg	Gly	Leu	Pro	Gly
10	U 1,				85					90		J	•		95	
	Thr	Ala	Gly	Leu	Pro	Gly	Met	Lys	Gly	His	Arg	Gly	Phe	Ser	Gly	Leu
15				100					105					110		
														_		
	Asp	Gly		Lys	Gly	Asp	Ala		Pro	Ala	Gly	Pro		Gly	Glu	Pro
20			115					120					125			
	G) v	Ser	Pro	Glv	Glu	Asn	Glv	Ala	Pro	Glv	Gln	Met	Gly	Pro	Arg	Gly
	07	130		,			135					140	•		-3	
25																
	Leu	Pro	Gly	Glu	Arg	Gly	Arg	Pro	Gly	Ala	Pro	Gly	Pro	Ala	Gly	Ala
30	145					150					155					160
	Arg	Gly	Asn	Asp	-	Ala	Thr	Gly	Ala	Ala	Gly	Pro	Pro	Gly		Thr
 35					165					170					175	
	Glv	Pro	Ala	Glv	Pro	Pro	Glv	Phe	Pro	Gly	Ala	Val	Gly	Ala	Lys	Gly
	 ,			180					185				•	190	•	•
40·																
	Glu	Ala	Gly	Pro	Gln	Gly	Pro	Arg	Gly	Ser	Glu	Gly	Pro	Gln	Gly	Val
			195					200					205			
45·				_			_		_					~~	_	_ •
	Arg			Pro	Gly	Pro		Gly	Pro	Ala	GIY	A1a 220	Ala	Gly	Pro	Ala
		210					215					220				
50	Gly	Asn	Pro	Gly	Ala	Asp	Gly	Gln	Pro	Gly	Ala	Lys	Gly	Ala	Asn	Gly
	225			_		230	•			-	235		J			240

5	Ala	Pro	Gly	Ile	Ala 245	Gly	Ala	Pro	Gly	Phe 250	Pro	Gly	Ala	Arg	Gly 255	Pro
	Ser	Gly	Pro	Gln	Gly	Pro	Gly	Gly	Pro	Pro	Gly	Pro	Lys	Gly	Asn	Ser
10				260					265					270		
15	Gly	Glu	Pro 275	Gly	Ala	Pro	Gly	Ser 280	Lys	Gly	Asp	Thr	Gly 285	Ala	Lys	Gly
:	Glu	Pro 290	Gly	Pro	Val	Gly	Val 295	Gln	Gly	Pro	Pro	Gly 300	Pro	Ala	Gly	Glu
20	Glu		Lys	Arg	Gly	Ala		Gly	Glu	Pro	Gly		Thr	Gly	Leu	Pro
25	305	•	•	_	-	310	-	•			315					320
20	Gly	Pro	Pro	Gly	Glu 325	Arg	Gly	Gly	Pro	Gly 330	Ser	Arg	Gly	Phe	Pro 335	Gly
30	Ala	Asp	Gly	Val	Ala	Gly	Pro	Lys	Gly 345	Pro	Ala	Gly	Glu	Arg 350	Gly	Ser
35	Pro	Gly	Pro	Ala	Gly	Pro	Lys	Gly		Pro	Gly	Glu	Ala		Arg	Pro
40		•	355		-		•	360			_		365		_	
	Gly	Glu 370	Ala	Gly	Leu	Pro	Gly 375	Ala	Lys	Gly	Leu	Thr 380	Gly	Ser	Pro	Gly
45	Ser 385		Gly	Pro	Asp	Gly 390	_	Thr	Gly	Pro	Pro 395	-	Pro	Ala	Gly	Gln 400
50	Asp	Gly	Arg	Pro	Gly		Pro	Gly	Pro	Pro	Gly	Ala	Arg	Gly	Gln 415	Ala

5	Gly	Val	Met	Gly	Phe	Pro	GJA	Pro	Lys	Gly	Ala	Ala	Gly	Glu	Pro	Gly
3				420					425					430		
	Lys	Ala	Gly	Glu	Arg	Gly	Val	Pro	Gly	Pro	Pro	Gly	Ala	Val	Gly	Pro
10	•		435			•		440	-				445		•	
2																
		61. 4	T	Asp	~1 11	~1	7 T n	G1	212	Gln.	G1v	Pro	Bro	G1v	Dwa	21-
4.5	AIA	_	гуя	Meb	GIA	GIU		GLY	MIG	GIII	Gry		PLO	GIY	PIO	ATA
15		450					455					460				
	Glγ	Pro	Ala	Gly	Glu	Arg	Gly	Glu	Gln	Gly		Ala	Gly	Ser	Pro	Gly
20	465					470					475					480
					•											
	Phe	Gln	Gly	Leu	Pro	Gly	Pro	Ala	Gly	Pro	Pro	GJA	Glu	Ala	Gly	Lys
					485					490					495	
25																
÷.	Pro	Gly	Glu	Gln	Gly	Val	Pro	Gly	Asp	Leu	Gly	Ala	Pro	Gly	Pro	Ser
•		-		500	•			•	505					510		
30	*															
g	~1·-	חות	7~~	Gl v	Gl v	7~~	G1 v	Dha	Pro	G] v	Glu	Ara	Glv	Va I	G] n	Gly
	GIY	AIG	-	GLY	GIU	AL 9	GLY		110	Gry	01 u	Æg	525	7 41	0111	GLY
0.5			515					520					745			
35·				_				_ '			_		_ •	_		
* *	Pro		Gly	Pro	Ala	Gly		Arg	Gly	Ala	Asn		Ala	Pro	Gly	Asn
		530					535					540				
40																
	Asp	Gly	Ala	Lys	Gly	Asp	Ala	Gly	Ala	Pro	Gly	Ala	Pro	Gly	Ser	Gln
	545					550					555					560
45										•						
45	Gly	Ala	Pro	Gly	Leu	Gln	Gly	Met	Pro	Gly	Glu	Arg	Gly	Ala	Ala	Gly
					565					570					575	
•																
50	Leu	Pro	Glv	Pro	Lvs	Glv	Asp	Ara	Glv	Asp	Ala	Glv	Pro	Lvs	Gly	Ala
:			4	580		2		3	585		•	-4		590		
									-03							

5	qaA	Gly	Ser	Pro	Gly	Lys	Asp	Gly	Val	Arg	Gly	Leu	Thr	Gly	Pro	Ile
			595					600					605			
	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Ala	Pro	Gly	Asp	Lys	Gly	Glu	Ser	Gly
10		610					615					620				
	Pro	Ser	Gly	Pro	Ala	Gly	Pro	Thr	Gly	Ala	Arg	Gly	Ala	Pro	Gly	qaA
15	625					630					635					640
	Arg	Gly	Glu	Pro	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Phe	Ala	Gly	Pro	Pro
20					645					650					655	
	Gly	Ala	Asp	Gly	Gln	Pro	Gly	Ala	Lys	Gly	Glu	Pro	Gly	Asp	Ala	Gly
25				660					665					670		
	Ala	Lys	_	Asp	Ala	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Pro	Ala	Gly	Pro
30			675					680					685			
	Pro	Gly	Pro	Ile	Gly	Asn	Val	Gly	Ala	Pro	Gly	Ala	Lys	Gly	Ala	Arg
25		690					695					700				
35	Gly	Ser	Ala	Gly	Pro	Pro	Gly	Ala	Thr	Gly	Phe	Pro	Gly	Ala	Ala	Gly
•	705					710					715					720
40	ሽ ን ርር	17s 1	alv	Pro	Pro	al v	Pro	Car	61.4	Zan	λla	Gl v	Dro	Dro	~1	D===
	Æg	V 0.1	GIY	210	725	GIŞ	PIO	961	GIY	730	ALG	GIY	PIO	PIO	735	PIO
45	Pro	Glv	Pro	Δla	Gly	Tava	G] v	G) v	G] v	Tva	Glv	Pro	7~~	Clv.	Gl.	Mp av
	110	GIJ	110	740	Cly		Giu	GIY	745		GIY	210	ALG	750	GIU	1111
50	Glv	Pro	Δla	Glv	Ara	Pro	Glv	Glu	l eV	Glv	Pro	Pro	G) v	Pro	Dro	Gly
	1		755	2	3			760		ung			765			31

. 5	Pro	Ala	Gly	Glu	Lys	Gly	Ser	Pro	Gly	Ala	Asp	Gly	Pro	Ala	Gly	Ala
		770					775					780				
	Dro	Glv	Thr.	Pro	Glv	Pro	G] n	G) v	Ile	٠.	Gly	Gln	n	C111	17-1	
10	785	Gly	****	710	GIY	790	GIII	GLY	116	AIG	795	GIII	ALY	GIÀ	val	800
																500
	Gly	Leu	Pro	Gly	Gln	Arg	Gly	Glu	Arg	Gly	Phe	Pro	Gly	Leu	Pro	Gly
. 15					805					810					815	
	Pro	Ser	Gly		Pro	Gly	Lys	Gln	Gly	Pro	Ser	Gly	Ala		Gly	Glu
20				820					825					830		
	Arg	Glv	Pro	Pro	Glv	Pro	Met	Glv	Pro	Pro	Glv	Leu	Ala	Glv	Pro	Pro
25	3		835					840					845	,		
· .	Gly	Glu	Ser	Gly	Arg	Glu	Gly	Ala	Pro	Ala	Ala	Glu	Gly	Ser	Pro	Gly
. 30		850					855					860				
	•	•	a 3	0	D	~1		•	0 3	•	-	a 1	41	-1		_
	865	Asp	GIY	ser	PTO	870	Ala	гÀв	Gly	Asp	875	GIA	GIU	Thr	GIA	880
35	003					0.0					0.5		ŕ			550
	Ala	Gly	Pro	Pro	Gly	Ala	Pro	Gly	Ala	Pro	Gly	Ala	Pro	Gly	Pro	Val
					885					890					895	
40																
	Gly	Pro	Ala		Lys	Ser	Gly	qaA	Arg	Gly	Glu	Thr	Gly		Ala	Gly
				900					905					910		
45	Pro	Ala	Glv	Pro	Val	Glv	Pro	Ala	Gly	Ala	Ara	Glv	Pro	Ala	Glv	Pro
			915		,	2		920			5	u-7	925		01	
50	Gln	Gly	Pro	Arg	Gly	Asp	Lys	Gly	Glu	Thr	Gly	Glu	Gln	Gly	qaA	Arg
		930					935					940				
						`										

		Gly	Ile	Lys	Gly	His	Arg	Gly	Phe	ser	Gly	Leu	Gln	Gly	Pro	Pro	Gly	
5		945					950					955					960	
											•							
		Pro	Pro	Gly	Ser		Gly	Glu	Gln	Gly		Ser	Gly	Ala	Ser	Gly	Pro	
: 10						965					970					975		
		λla	Glv	Pro	Ara	Glv	Pro	Pro	Glv	Ser	Ala	Glv	Ala	Pro	Glv	Lys	Asp	
15		7,14	017		980	1			,	985					990	-1-		
		Gly	Leu	Asn	Gly	Leu	Pro	Gly	Pro	Ile	Gly	Pro	Pro	Gly	Pro	Arg	Gly	
20				995					100	0				100	5			
						_			_	_	_			_	_			
		Arg			Asp	Ala	Gly			Gly	Pro	Pro			Pro	Gly	Pro	•
-25			101	U			•	101	>				102	U				
		Pro	Gly	Pro	Pro	Gly	Pro	Pro	Ser	Ala	Gly	Phe	Asp	Phe	Ser	Phe	Leu	
		102	_				103					103					1040	
30																		
		Pro	Gln	Pro	Pro	Gln	Glu	Lys	Ala	His	Asp	Gly	Gly	Arg	Tyr	Tyr	Arg	
						104	5			٠	105	0				105	5	
35		21.														•		
		Ala																
40																		
40	(2)	INFO	RMAT	ION	FOR	SEQ	ID N	0:17	:									
45		(i)	_		E CH													
				•	ngth Pe :				cids									
			•	•	PE: RAND				le									
50					POLO			_										
		(ii)	MOL	ECUL	E TY	PE:	pept	ide										

_	(1X) FEATURE:			
. 5	(A) NAME/KEY: Region			
	(B) LOCATION: 12	•		
	(D) OTHER INFORMATION	: /note= "Ami	no acid seque	nce for
10	glutathione S-transferase*			
	(ix) FEATURE:			
15	(A) NAME/KEY: Region			
•	(B) LOCATION: 1920			
	(D) OTHER INFORMATION	: /note= "338	repeats of t	he
20	following triplet Gly-X-y when	rein about 35	t of the X and	d Y
	positions are occupied by pro-	line and 4-hy	droxyproline.	и
25	(xi) SEQUENCE DESCRIPTION:	SEQ ID NO:17:		
	Xaa Met Gln Leu Ser Tyr Gl	y Tyr Asp Glu	Lys Ser Thr	Gly Gly Ile
30	1 5	10		15
	Ser Val Pro Xaa Ser Ala Gl	y Phe Asp Phe	Ser Phe Leu	Pro Gln Pro
35	20	25		30
	Pro Gln Glu Lys Ala His Asp	o Gly Gly Arg	Tyr Tyr Arg	Ala
40	35	40	45	
3				
	(2) INFORMATION FOR SEQ ID NO:1	8:		
45				
45	(i) SEQUENCE CHARACTERISTIC			
	(A) LENGTH: 31 amino	acids		
	(B) TYPE: amino acid			
50	(C) STRANDEDNESS: sin	gle		
	(D) TOPOLOGY: unknown			
	(ii) MOLECULE TYPE: peptide			

	(ix) FEATURE:
5	(A) NAME/KEY: Region .
	(B) LOCATION: 12
	(D) OTHER INFORMATION: /note= "Amino acid sequence for
10	glutathione S-transferase."
1	
	(ix) FEATURE:
15	(A) NAME/KEY: Region
	(B) LOCATION: 45
	(D) OTHER INFORMATION: /note= "338 repeats of the
20	following triplet Gly-X-Y wherein about 35% of the X and Y
	positions are occupied by proline and 4-hydroxyproline. "
25	(wi) CENTENCE DESCRIPTION, SEC ID NO.10.
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:
	Xaa Met Gly Xaa Tyr Ser Ala Gly Phe Asp Phe Ser Phe Leu Pro Gln
30	1 5 10 15
	Pro Pro Gln Glu Lys Ala His Asp Gly Gly Arg Tyr Tyr Arg Ala
35	20 25 30
	(2) INFORMATION FOR SEQ ID NO:19:
40 [.]	
	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 3171 base pairs
4 5	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
	(D) TOPOLOGY: linear
50	(ii) MOLECULE TYPE: DNA (genomic)
	(11) transcomments of the state of the s

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

٠	5

10

20

25

30

35 .

40

45

50

CAGCTGAGCT ATGGCTATGA TGAAAAAAGC ACCGGCGGCA TCAGCGTGCC GGGCCCGATG 60 GGTCCGAGCG GCCCTCGTGG CCTGCCGGGC CCGCCAGGTG CGCCCGGTCC GCAGGGCTTT 120 CAGGGTCCGC CGGGCGAACC GGGCGAACCT GGTGCGAGCG GCCCGATGGG CCCGCGCGGC 180 CCGCCGGGTC CGCCAGGCAA AAACGGCGAT GATGGCGAAG CGGGCAAACC GGGACGTCCG 240 GGTGAACGTG GCCCCCGGG CCCGCAGGGC GCGCGCGGAC TGCCGGGTAC TGCGGGACTG 300 CCGGGCATGA AAGGCCACCG CGGTTTCTCT GGTCTGGATG GTGCGAAAGG TGATGCGGGT CCGGCGGGTC CGAAAGGTGA GCCGGGCAGC CCGGGCGAAA ACGGCGCGCC GGGTCAGATG 420 GGCCCGCGTG GCCTGCCTGG TGAACGCGGT CGCCCGGGCG CCCCGGGCCC AGCTGGCGCA 480 CGTGGCAACG ATGGTGCGAC CGGTGCGGCC GGTCCACCGG GCCCGACGGG CCCGGCGGGT 540 CCCCGGGCT TTCCGGGTGC GGTGGGTGCG AAAGGCGAAG CAGGTCCGCA GGGGCCGCGC 600 GGGAGCGAGG GTCCTCAGGG CGTTCGTGGT GAACCGGGCC CGCCGGGCCC GGCGGGTGCG 660 GCGGGCCCGG CTGGTAACCC TGGCGCGGAC GGTCAGCCAG GTGCGAAAGG TGCCAACGGC 720 GCGCCGGGTA TTGCAGGTGC ACCGGGCTTC CCGGGTGCCC GCGGCCCGTC CGGCCCGCAG 780 GGCCCGGGCG GCCCGCCCGG CCCGAAAGGG AACAGCGGTG AACCGGGTGC GCCAGGCAGC 840 900 AAAGGCGACA CCGGTGCGAA AGGTGAACCG GGCCCAGTGG GTGTTCAAGG CCCGCCGGGC 960 CCGGCGGGCG AGGAAGGCAA ACGCGGTGCT CGCGGTGAAC CGGGCCCGAC CGGCCTGCCT

55

j

.

 $/\kappa^{\kappa}$

40:

•	GGCCCGCCGG	GAGAACGTGG	TGGCCCGGGT	AGCCGCGGTT	TTCCGGGCGC	GGATGGTGTG	1020
•	GCGGGCCCGA	AAGGTCCGGC	GGGTGAACGT	GGTAGCCCĢG	GCCCGGCGGG	CCCAAAAGGC	1080
•	AGCCCGGGCG	AGGCAGGACG	TCCGGGTGAA	GCGGGTCTCC	CGGGCGCCAA	AGGTCTGACC	1140
	GGCTCTCCGG	GCAGCCCGGG	TCCGGATGGC	AAAACGGGCC	CGCCTGGTCC	GGCCGGCCAG	1200
	GATGGTCGCC	CGGGCCCGCC	GGGCCCGCCG	GGTGCCCGTG	GTCAGGCGGG	TGTCATGGGC	1260
	TTTCCAGGCC	CCAAAGGTGC	GGCGGGTGAA	CCGGGCAAAG	CGGGCGAACG	CGGTGTCCCG	1320
	GGTCCGCCGG	GCGCTGTCGG	GCCGGCGGGC	AAAGATGGCG	AAGCGGGCGC	GCAAGGCCCG	1380
	CCGGGACCAG	CGGGTCCGGC	GGGCGAGCGC	GGTGAACAGG	GCCCGGCAGG	CAGCCCGGGT	1440
	TTCCAGGGTC	TGCCGGGCCC	TGCGGGTCCA	CCGGGTGAAG	CGGGCAAACC	GGGGGAACAA	1500
	GGTGTGCCGG	GCGACCTGGG	CGCCCCAGGC	CCGAGCGGCG	CGCGCGCGA	ACGCGGTTTC	1560
	CCGGGCGAAC	GTGGTGTGCA	GGGCCGCCC	GCCCGGCTG	GTCCGCGCGG	CGCCAACGGC	1620
	GCGCCGGGCA	ACGATGGTGC	GAAAGGTGAT	GCGGGTGCCC	CAGGTGCGCC	GGGCAGCCAG	1680
	GGCGCCCGG	GGCTGCAAGG	CATGCCGGGT	GAACGTGGTG	CCGCGGGTCT	ACCGGGTCCG	1740
	AAAGGCGACC	GCGGTGATGC	GGGTCCAAAA	GGTGCGGATG	GCTCCCCTGG	CAAAGATGGC	1800
	GTTCGTGGTC	TGACCGGCCC	GATCGGCCCG	ccggcccgg	CAGGTGCCCC	GGGTGACAAA	1860
	GGTGAAAGCG	GTCCGAGCGG	CCCAGCGGGC	CCCACTGGTG	CGCGTGGTGC	CCCGGGCGAC	1920
	CGTGGTGAAC	CGGGTCCGCC	GGGCCCGGCG	GGCTTTGCGG	GCCCGCCAGG	CGCTGACGGC	1980

.!

ţ,

20.

30.

-.

CAGCCGGGTG	CGAAAGGCGA	ACCGGGGGAT	GCGGGTGCTA	AAGGCGACGC	GGGTCCGCCG	2040
GGCCCTGCCG	GCCCGGCGGG	CCCGCCAGGC	CCGATTGGCA	ACGTGGGTGC	GCCGGGTGCC	2100
AAAGGTGCGC	GCGGCAGCGC	TGGTCCGCCG	GGCGCGACCG	GTTTCCCCGG	TGCGGCGGG	2160
CGCGTGGGTC	CGCCAGGCCC	GAGCGGTAAC	GCGGGTCCGC	CAGGTCCGCC	TGGCCCGGCT	2220
GGCAAAGAGG	GCGGCAAAGG	TCCGCGTGGT	GAAACCGGCC	CTGCGGGACG	TCCAGGTGAA	2280
GTGGGTCCGC	CGGGCCCGCC	GGGCCCGGCG	GGCGAAAAAG	GTAGCCCGGG	TGCGGATGGT	2340
CCCGCCGGTG	CGCCAGGCAC	GCCGGGTCCG	CAAGGTATCG	CTGGCCAGCG	TGGTGTCGTC	2400
GGGCTGCCGG	GTCAGCGCGG	CGAACGCGGC	TTTCCGGGTC	TGCCGGGCCC	GAGCGGTGAG	2460
CCGGGCAAAC	AGGGTCCATC	TGGCGCGAGC	GGTGAACGTG	GCCCGCCGGG	TCCCATGGGC	2520
CCGCCGGGTC	TGGCGGGCCC	TCCGGGTGAA	AGCGGTCGTG	AAGGCGCGCC	GGGTGCCGAA	2580
GGCAGCCCAG	GCCGCGACGG	TAGCCCGGGG	GCCAAAGGGG	ATCGTGGTGA	AACCGGCCCG	2640
GCGGGCCCCC	CGGGTGCACC	GGGCGCCG	GGTGCCCCAG	GCCCGGTGGG	CCCGGCGGGC	2700
AAAAGCGGTG	ATCGTGGTGA	GACCGGTCCG	GCGGGCCCGG	CCGGTCCGGT	GGGCCCAGCG	2760
GGCGCCCGTG	GCCCGGCCGG	TCCGCAGGGC	CCGCGGGGTG	ACAAAGGTGA	AACGGGCGAA	2820
CAGGGCGACC	GTGGCATTAA	AGGCCACCGT	GGCTTCAGCG	GCCTGCAGGG	TCCACCGGGC	2880
CCGCCGGGCA	GTCCGGGTGA	ACAGGGTCCG	TCCGGAGCCA	GCGGGCCGGC	GGGCCCACGC	2940
GGTCCGCCGG	GCAGCGCGGG	CGCGCCGGGC	AAAGACGGTC	TGAACGGTCT	GCCGGGCCCG	3000

45·

.

ATCGGCCCG	C CG	GGCC	CACG	CGG	CCGC	ACC	GGTG	ATGC	GG G	TCCG	GTGG	G TC	cccc	GGGC	!	3060
CCGCCGGGC	C CG	CCAG	GCCC	GCC	GGGA	.CCG	CCGA	.GCGC	GG G	TTTC	GACT	T CA	GCTI	CCTG	}	3120
CCGCAGCCG	C CG	CAGG	AGAA	AGC	GCAC	GAC	GGCG	GTCG	CT A	CTAC	CGTG	C G				3171
(2) INFOR	ITAM	ON F	OR S	EQ I	D NO	:20:										
(i)	SEQU	ENCE	CHA	RACI	ERIS	TICS	ß:									
	(A)	LEN	GTH:	105	7 am	ino	acid	ls								
	(B)	TYP	E: a	mino	aci	.d										
•	(C)	STR	ANDE	DNES	S: 8	ingl	.e									
	(D)	TOP	OLOG	¥:υ	ınkno	wn										•
(ii) (xi)	٠						II QE) NO:	20:							
Gln	Leu	Ser	Tyr	Gly	Tyr	Asp	Glu	Lys	Ser	Thr	Gly	Gly	Ile	Ser	Val	
1				5					10					15		
Pro	Gly	Pro	Met	Gly	Pro	Ser	Gly	Pro	Arg	Gly	Leu	Pro	Gly	Pro	Pro	
			20					25			•		30			
Gly	Ala	Pro	Gly	Pro	Gln	Gly	Phe	Gln	Gly	Pro	Pro	Gly	Glu	Pro	Gly	
		35			5		40					45	•			
Glu	Pro	Gly	Ala	Ser	Gly	Pro	Met	Gly	Pro	Arg	Gly	Pro	Pro	Gly	Pro	
	50					55					60					

5	Pro	Gly	Lys	Asn	Gly	Asp	Asp	Gly	Glu	Ala	Gly	Lys	Pro	Gly	Arg	Pro
3	65					70					75					80
•	Gly	Glu	Arg	Gly	Pro	Pro	Gly	Pro	Gln	Gly	Ala	Arg	Gly	Leu	Pro	Glv
10	_		_		85		-			90			•		95	
										20					75	
	ml		a 1	T	D	~ 1	14 - A	•	~1	***	• • • • •	a 3	n t	_		
	THE	Ala	GIA		PIO	GIÀ	met	гЛа		HIS	Arg	GIY	РЛЕ	Ser	Gly	Leu
15				100					105					110		
,	Asp	Gly	Ala	Lys	Gly	Asp	Ala	Gly	Pro	Ala	Gly	Pro	Lys	Gly	Glu	Pro
20			115					120					125			
	Gly	Ser	Pro	Gly	Glu	Asn	Gly	Ala	Pro	Gly	Gln	Met	Gly	Pro	Arg	Gly
		130					135					140			_	_
25																
	Leu	Pro	Glv	Glu	Ara	Glv	Ara	Pro	Glv	Ala	Pro	Glv	Pro	Ala	Glv	Δ1 =
÷	145		2		3	150	3		1		155	,			Cly	
30						130					175					160
		67	•	•	a 1							_	_			
	Arg	GIŢ	ASI	Asp		ATA	Thr	GIY	Ala		GIÅ	Pro	Pro	Gly	Pro	Thr
					165					170					175	
35																
	Gly	Pro	Ala	Gly	Pro	Pro	Gly	Phe	Pro	Gly	Ala	Val	Gly	Ala	Lys	Gly
÷.				180					185					190		
40 [:]																
	Glu	Ala	Gly	Pro	Gln	Gly	Pro	Arg	Gly	Ser	Glu	Gly	Pro	Gln	Gly	Val
			195					200					205			
45	Arg	Gly	Glu	Pro	Gly	Pro	Pro	Gly	Pro	Ala	Glv	Ala	Ala	Gly	Pro	Ala
		210			_		215		-			220		,		
												V				
50	Gl v	Aer	Dro	Gl v	λ 1 -	X are	~1. -	~1 =	Dwa	al. .	83 m.	T	~1. -	Ala	>	~ 1
		van	210	Gry	vra		GTÅ	GIU	PĽO	стλ		пÀв	GIA	AIA	ASU	
	225					230					235					240

55

.

	Ala	Pro	Gly	Ile	Ala	Gly	Ala	Pro	Gly	Phe	Pro	Gly	Ala	Arg	Gly	Pro
5					245					250					255	
•	Ser	Gly	Pro	Gln	Gly	Pro	Gly	Gly	Pro	Pro	Gly	Pro	Lys	Gly	Asn	Ser
10		_		260					265					270		
:	Gly	Glu	Pro	Glv	Δla	Pro	Glv	Ser	Lva	Glv	Asp	Thr	Glv	Ala	Ive	Gly
	GIY	GIU	275	O.J			01,	280	-,-	,			285		_,,	O.J
15			273					200					203			
		_	~ 3	D	*** 1	a1	*** 3	a 1-	a 1	D	D===	~ 3	Dwa	71 -	01	~1
.F	GIU		GIY	PIO	vaı	GIA		GIN	GIY	PIO	PIO		PIO	AId	GIÀ	Glu
20		290					295					300				
														_		
٠.	Glu	Gly	Lys	Arg	Gly	Ala	Arg	Gly	Glu	Pro	Gly	Pro	Thr	Gly	Leu	Pro
	305					310					315					320
25																
	Gly	Pro	Pro	Gly	Glu	Arg	Gly	Gly	Pro	Gly	Ser	Arg	Gly	Phe	Pro	Gly
					325					330					335	
30																
	Ala	Asp	Gly	Val	Ala	Gly	Pro	Lys	Gly	Pro	Ala	Gly	Glu	Arg	Gly	Ser
				340					345					350		
25																
35	Pro	Gly	Pro	Ala	Gly	Pro	Lys	Gly	Ser	Pro	Gly	Glu	Ala	Gly	Arg	Pro
« :		_	355					360					365			
•																
40	Glv	Glu	Ala	Glv	Leu	Pro	Glv	Ala	Lvs	Glv	Leu	Thr	Gly	Ser	Pro	Gly
	0-7	370		,			375					380	•			
		3.0														
45	Com	Bro	G1 v	Dro) an	Glv	Tara	Thr	G1v	Pro	Dro	Glv	Dro	λla	Glv	Gln
			GIY	FIU	veb		Lyb	4111	Gry	710	395			nza	017	400
	385					390					273					~UU
•	_		_	_		_	_		_	_	 3		•		~ 1	
50	Asp	Gly	Arg	Pro	_	Pro	Pro	Gly	Pro		GIA	Ala	Arg	GTÅ		Ala
					405					410					415	

	Gly	Val	Met	Gly	Phe	Pro	Gly	Pro	Lys	Gly	Ala	Ala	Gly	Glu	Pro	Gly
5				420					425					430		
	Lys	Ala	Gly	Glu	Arg	Gly	Val	Pro	Gly	Pro	Pro	Gly	Ala	Val	Gly	Pro
10			435			_		440	-			_	445		-	
	Ala		Lys	Asp	Gly	Glu	Ala	Gly	Ala	Gln	Gly		Pro	Gly	Pro	Ala
15		450					455					460				
	Glv	Pro	Ala	Glv	Glu	Arg	Glv	Glu	Gln	Glv	Pro	Ala	Gly	Ser	Pro	g) v
	465			•		470					475					480
20																
	Phe	Gln	Gly	Leu	Pro	Gly	Pro	Ala	Gly	Pro	Pro	Gly	Glu	Ala	Gly	Lys
25					485					490					495	
	D=0	C) v	<i>a</i> 1	al n	C1	17-1	D===	~1··	200	T 0	~1. .	N3 -	Pro	~1	D	.
	PIO	GIY	GIU	500	GIĀ	val	PIO	GIĀ	505	Leu	GIY	Ala	PIO	510	Pro	ser
30																
	Gly	Ala	Arg	Gly	Glu	Arg	Gly	Phe	Pro	Gly	Glu	Arg	Gly	Val	Gln	Gly
44			515					520					525			
35	_	_		_				_			_					
	Pro	Pro 530	Gly	Pro	Ala	Gly	Pro 535	Arg	Gly	Ala	Asn	Gly 540	Ala	Pro	Gly	Asn
,		330					333					340				
40	Asp	Gly	Ala	Lys	Gly	Asp	Ala	Gly	Ala	Pro	Gly	Ala	Pro	Gly	Ser	Gln
	545					550					555					560
45																
45	Gly	Ala	Pro	Gly		Gln	Gly	Met	Pro	_	Glu	Arg	Gly	Ala		Gly
					565					570					575	
50	Leu	Pro	Gly	Pro	Lys	Gly	Asp	Arg	Gly	Asp	Ala	Gly	Pro	Lys	Gly	Ala
							_		_	_		-			-	
				580					585					590		

Asp	Gly		Pro	Gly	Lys	Asp ·		Val	Arg	Gly	Leu		Gly	Pro	Ile
		575					600					605			
Gly		Pro	Gly	Pro	Ala		Ala	Pro	Gly	Asp		Gly	Glu	Ser	Gly
	610					615				ı	620				
	Ser	Gly	Pro	Ala		Pro	Thr	Gly	Ala	Arg	Gly	Ala	Pro	Gly	Asp
625					630					635					640
Arg	Gly	Glu	Pro	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Phe	Ala	Gly	Pro	Pro
				645					650					655	
Gly	Ala	Asp	Gly	Gln	Pro	Gly	Ala	Lys	Gly	Glu	Pro	Gly	Asp	Ala	Gly
			660					665					670		
Ala	Lys	Gly	Asp	Ala	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Pro	Ala	Gly	Pro
		675					680					685			
Pro	Gly	Pro	Ile	Gly	Asn	Val	Gly	Ala	Pro	Gly	Ala	Lys	Gly	Ala	Arg
	690					695					700				
Gly	Ser	Ala	Gly	Pro	Pro	Gly	Ala	Thr	Gly	Phe	Pro	Gly	Ala	Ala	Gly
705					710					715					720
Arg	Val	Gly	Pro	Pro	Gly	Pro	Ser	Gly	Asn	Ala	Gly	Pro	Pro	Gly	Pro
				725					730					735	
Pro	Gly	Pro	Ala	Gly	Lys	Glu	Gly	Gly	Lys	Gly	Pro	Arg	Gly	Glu	Thr
			740					745					7 50		
Gly	Pro	Ala	Gly	Arg	Pro	Gly	Glu	Val	Gly	Pro	Pro	Gly	Pro	Pro	Gly
		755					760					765			
	Gly Pro 625 Arg Gly 705 Arg	Gly Pro 610 Pro Ser 625 Arg Gly Gly Ala Ala Lys Pro Gly 690 Gly Ser 705 Arg Val	Gly Pro Pro 610 Arg Gly Glu Gly Ala Asp Ala Lys Gly 675 Pro Gly Pro 690 Gly Ser Ala 705 Arg Val Gly Pro Gly Pro	S95 Gly Pro Gly Fro Gly Glu Pro Gly Glu Glu Gly Glu Gly Gro Gly Gro Gly Gro Gly Fro Gly Gly Fro Gly Fro Gly Fro	S95	Gly Pro Pro Gly Pro Ala Pro Ser Gly Pro Ala Gly 630 Arg Gly Gly Pro Gly Pro 645 Ala Lys Gly Asp Ala Gly Pro Asp Ala Gly Pro Asp Ala Gly Asp Ala A	Gly Pro Pro Gly Pro Ala Gly 615 Pro Ser Gly Pro Ala Gly Pro 615 Pro Gly Pro Ala Gly Pro Pro Pro Pro Pro Pro Gly Pro Pro Gly Pro Pro Pro Gly Pro Pro Pro Gly Pro Pro Pro Gly Pro P	Gly Pro Pro Gly Pro Ala Gly Ala Gly Ala Ala Gly Pro Thr Gly Pro Ala Gly Pro Thr Gly Pro Thr Gly Pro Thr Gly Pro P	Gly Pro Pro Pro Gly Pro Ala Gly Ala Pro Gly Pro 610 Pro Gly Pro Gl	Gly Pro Pro Pro Gly Pro Gly 610 Gly Pro Gly 610 Ala Gly Ala Pro Gly 615 Pro Gly Ala Gly Ala Gly Pro Thr Gly Ala 630 Ala Gly Pro Thr Gly Ala 630 Ala Gly Pro Thr Gly Ala 650 Ala Gly Pro Gly Fro Gly 650 Ala 630 Pro Gly Pro Ala 650 Ala 630 Pro Gly Pro Ala 650 Ala 645 Pro Gly Pro Gly 650 Ala 650 Ala 650 Pro Gly 650 Ala 650 Pro Gly 650 Ala 650 Pro Gly 660 Pro Gly 660 Pro Gly 660 Pro Gly 700 Pro 700<	See See	Gly Pro Gly Pro Ala Gly Ala Pro Gly Pro Ey8 Gly Pro Fro Gly Pro F	S95	S95	Gly Pro Pro Gly Pro Ala Gly Pro Ala Gly Ala Pro Gly Asp Lys Gly Glu Ser 610

5	Pro	Ala	Gly	Glu	Lys	Gly	Ser	Pro	Gly	Ala	Asp	Gly	Pro	Ala	Gly	Ala
		770					775					780				
10	Pro	Gly	Thr	Pro	Gly	Pro	Gln	Gly	Ile	Ala	Gly	Gln	Arg	Gly	Val	Val
	785					790					795					800
<u> </u>	Gly	Leu	Pro	Gly	Gln	Arg	Gly	Glu	Arg	Gly	Phe	Pro	Gly	Leu	Pro	Gly
15					805					810					815	
ī	Pro	Ser	Gly	Glu	Pro	Gly	Lys	Gln	Gly	Pro	Ser	Gly	Ala	Ser	Gly	Glu
20				820					825					830		
:																
	Arg	Gly	Pro	Pro	Gly	Pro	Met	Gly	Pro	Pro	Gly	Leu	Ala	Gly	Pro	Pro
.· 25			835					840					845			
	Gly	Glu	Ser	Gly	Arg	Glu	Gly	Ala	Pro	Gly	Ala	Glu	Gly	Ser	Pro	Gly
20		850					855					860				
30																
	Arg	Asp	Gly	Ser	Pro	Gly	Ala	Lys	Gly	Asp	Arg	Gly	Glu	Thr	Gly	Pro
	865					870					875					880
35																
	Ala	Gly	Pro	Pro	Gly	Ala	Pro	Gly	Ala	Pro	Gly	Ala	Pro	Gly	Pro	Val
D.					885					890					895	
40																
	Gly	Pro	Ala	Gly	Lys	Ser	Gly	Asp	Arg	Gly	Glu	Thr	Gly	Pro	Ala	Gly
				900					905					910		
45 [°]																
-	Pro	Ala	Gly	Pro	Val	Gly	Pro	Ala	Gly	Ala	Arg	Gly	Pro	Ala	Gly	Pro
			915					920					925			
50																
50	Gln	Gly	Pro	Arg	Gly	Asp	Lys	Gly	Glu	Thr	Gly	Glu	Gln	Gly	Asp	Arg
		930					935					940				

.t . 5		Gly 945	Ile	Lys	Gly	His	Arg 950	Gly	Phe	Ser	Gly	Leu 955	Gln	Gly	Pro	Pro	Gly 960
10		Pro	Pro	Gly	Ser	Pro 965	Gly	Glu	Gln	Gly	Pro 970	Ser	Gly	Ala	Ser	Gly 975	Pro
15		Ala	Gly	Pro	Arg 980	Gly	Pro	Pro	Gly	Ser 985	Ala	Gly	Ala	Pro	Gly 990	Lys	Asp
20		Gly	Leu	Asn 995	Gly	Leu	Pro	Gly	Pro 1000		Gly	Pro	Pro	Gly 100!		Arg	Gly
25		Arg	Thr 1010	_	Ąsp	Ala	Gly	Pro 1015		Gly	Pro	Pro	Gly 1020	Pro	Pro	Gly	Pro
30		Pro 1025	_	Pro	Pro	Gly	2ro 1030		Ser	Ala	Gly	Phe 1035	_	Phe	Ser	Phe	Leu 1040
-: 35		Pro	Gln	Pro	Pro	Gln 1045		Lys	Ala	His	Asp		Gly	Arg	Tyr	Tyr 1055	-
		Ala			-												
40	(2)	INFO	TAMS	ION I	FOR S	SEQ I	D NO):21:	:								
45		(i)	(A)	LEI	E CHU NGTH PE: 1	: 79	base	e pa									•
50					RANDI POLO			_	Le								
		(ii)	MOLI	CUL	E TY	PE: d	DNA										

4

55.

5 .	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:	
	GGAATTCATG CAGCTGAGCT ATGGCTATGA TGAAAAAAGC ACCGGCGGCA TCAGCGTGCC	60
	GGGCCCGATG GGTCCGAGC	79
15.	(2) INFORMATION FOR SEQ ID NO:22:	
20	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 75 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
30	(ii) MOLECULE TYPE: cDNA	
35	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:	
	GGCCCGGGCT ACCCAGGCTC GCCGGGCGCA CCGGACGGCC CGGGCGGTCC AGCGGGGCCA	60
40	GCATTATTCG AACCC	75
•	(2) INFORMATION FOR SEQ ID NO:23:	
45 . 50	(i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 81 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	

5	(ii) MOLECULE TYPE: cDNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:	
10	GGAATTCCGG GTCCGCAGGG CTTTCAGGGT CCGCCGGGCG AACCTGGTGC GAGCGGCCCG	60
••	ATGGGCCCGC GCGGCCCGCC C	81
	(2) INFORMATION FOR SEQ ID NO:24:	
	(i) SEQUENCE CHARACTERISTICS:	
20	(A) LENGTH: 87 base pairs	
	(B) TYPE: nucleic acid	•
	(C) STRANDEDNESS: single	
25	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: cDNA	
30		
:		
35	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:	
42	TACCCGGGCG CGCCGGGCGG CCCAGGCGGT CCGTTTTTGC CGCTACTACC GTTCGCCCGT	60
40	TTGGCCCTGC AGGCATTATT CGAACCC	87
	(2) INFORMATION FOR SEQ ID NO:25:	
45	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 111 base pairs	
	(B) TYPE: nucleic acid	
50 [.]	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	

5	(ii) MOLECULE TYPE: cDNA	
•	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:	
10	CAGCTGAGCT ATGGCTATGA TGAAAAAAGC ACCGGCGGCA TCAGCGTGCC GGGCCCGATG	60
15	GGTCCGAGCG GCCCTCGTGG CCTGCCGGGC CCGCCAGGTG CGCCCGGTCC G	111
•	(2) INFORMATION FOR SEQ ID NO:26:	
20	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 37 amino acids	
	(B) TYPE: amino acid	
•	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: unknown	
30	(ii) MOLECULE TYPE: peptide	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:	
35	Gln Leu Ser Tyr Gly Tyr Asp Glu Lys Ser Thr Gly Gly Ile Ser Val	
	1 5 10 15	
40	Pro Gly Pro Met Gly Pro Ser Gly Pro Arg Gly Leu Pro Gly Pro Pro	
40	20 25 . 30	
	Gly Ala Pro Gly Pro	
45·	35	
•		
	(2) INFORMATION FOR SEQ ID NO:27:	
50		
•	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 240 base pairs	
55	(B) TYPE: nucleic acid	

5	(C) STRANDEDNESS: Bingle	
	(D) TOPOLOGY: linear	
;		
10	(ii) MOLECULE TYPE: cDNA	
; ;*	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:	
15	CAGCTGAGCT ATGGCTATGA TGAAAAAAGC ACCGGCGGCA TCAGCGTGCC GGGCCCGATG	60
20	GGTCCGAGCG GCCCTCGTGG CCTGCCGGGC CCGCCAGGTG CGCCCGGTCC GCAGGGCTTT	120
	CAGGGTCCGC CGGGCGAACCT GGTGCGAGCG GCCCGATGGG CCCGCGCGCG	180
25	CCGCCGGGTC CGCCAGGCAA AAACGGCGAT GATGGCGAAG CGGGCAAACC GGGACGTCCG	240

30	(2) INFORMATION FOR SEQ ID NO:28:	
	(i) SEQUENCE CHARACTERISTICS:	
**.	(A) LENGTH: 80 amino acids	
35·	(B) TYPE: amino acid	
	(C) STRANDEDNESS: single	
۲,	(D) TOPOLOGY: unknown	
40	• • • • • • • • • • • • • • • • • • • •	
	(ii) MOLECULE TYPE: peptide	
45	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:	
	Gln Leu Ser Tyr Gly Tyr Asp Glu Lys Ser Thr Gly Gly Ile Ser Val	
	1 5 10 15	
50		
	Pro Gly Pro Met Gly Pro Ser Gly Pro Arg Gly Leu Pro Gly Pro Pro	
	20 25 30	

5	Gly	Ala	Pro	Gly	Pro	Gln	Gly	Phe	Gln	Gly	Pro	Pro	Gly	Glu	Pro	Gly	
3			35					40					45				
. 10	Glu		Gly	Ala	Ser	Gly		Met	Gly	Pro	Arg		Pro	Pro	Gly	Pro	
		50					55					60					
.1		Gly	Lys	Asn	Gly		Asp	Gly	Glu	Ala		Lys	Pro	Gly	Arg		
15	. 65					70					75					80	
20	(2) INFO	RMAT]	ION I	FOR S	SEQ :	ID NO	0:29	:									
٠	(i)	SEQU	JENCI	E CHA	ARAC:	rer i s	STICS	S :									•
25		(B)	TYI	NGTH:	nucle	eic a	acid		3								
				RANDI POLOC			_	Le									
30		,_,															
	(ii)	MOLI	ECULI	E TYI	PE: o	DNA											
35	(xi)	SEQ	JENCI	E DES	SCRII	PTIO	N: SI	EQ II	0 ио:	29:							
\$	CAGTATGA'	TG G/	AAAA	GGAGT	TGO	SACT	rggc	CCT	GACC	'AA 2	rege	CTTA	AT GO	GAC	CTAGA		60
40	GGCCCACC"	rg gi	rgca	GCTGC	AG(cccı	AGGC	CCT	CAAGG	TT 1	CCA	AGGA	CC TO	GCTG(STGAG	}	120
45	CCTGGTGA	AC C	rggt(CAAAC	TG(TCC	rgca	GGT	CTCG	TG (TCC/	GCT	G C	CCTC	CTGGC	:	180
	AAGGCTGG'	TG AJ	AGATO	GTC#	CCC	CTGG <i>i</i>	LAAA	ccc	GACG	AC (CTGGT	(DAD?	AG AC	GAG1	TGT		240
50	GGACCACA	GG G1	rgct(CGTGC	TT	rccci	rgga	ACTO	CTGG	AC 1	TCC	rggen	TT C	\AAG(CATI	•	300
	AGGGGACA	CA AT	rggt(CTGG	TGO	SATTO	BAAG	GGA	CAGCC	CG C	TGC	CCT	G TO	STGA	AGGGT		360
55																	

5	GAACCTGGTG	CCCCTGGTGA	AAATGGAACT	CCAGGTCAAA	CAGGAGCCCG	TGGGCTTCCT	420
	GGTGAGAGAG	GACGTGTTGG	TGCCCCTGGC	CCAGCTGGTG	CCCGTGGCAG	TGATGGAAGT	480
10	GTGGGTCCCG	TGGGTCCTGC	TGGTCCCATT	GGGTCTGCTG	GCCCTCCAGG	CTTCCCAGGT	540
15	GCCCCTGGCC	CCAAGGGTGA	AATTGGAGCT	GTTGGTAACG	CTGGTCCTGC	TGGTCCCGCC	600
;	GGTCCCCGTG	GTGAAGTGGG	TCTTCCAGGC	CTCTCCGGCC	CCGTTGGACC	TCCTGGTAAT	660
20	CCTGGAGCAA	ACGGCCTTAC	TGGTGCCAAG	GGTGCTGCTG	GCCTTCCCGG	CGTTGCTGGG	720
	GCTCCCGGCC	TCCCTGGACC	CCGCGGTATT	CCTGGCCCTG	TTGGTGCTGC	CGGTGCTACT	780
25	GGTGCCAGAG	GACTTGTTGG	TGAGCCTGGT	CCAGCTGGCT	CCAAAGGAGA	GAGCGGTAAC	840
30	AAGGGTGAGC	CCGGCTCTGC	TGGGCCCCAA	GGTCCTCCTG	GTCCCAGTGG	TGAAGAAGGA	900
	AAGAGAGGCC	CTAATGGGGA	AGCTGGATCT	GCCGGCCCTC	CAGGACCTCC	TGGGCTGAGA	960
35 ⁻	GGTAGTCCTG	GTTCTCGTGG	TCTTCCTGGA	GCTGATGGCA	GAGCTGGCGT	CATGGGCCCT	1020
	CCTGGTAGTC	GTGGTGCAAG	TGGCCCTGCT	GGAGTCCGAG	GACCTAATGG	AGATGCTGGT	1080
40	CGCCCTGGGG	AGCCTGGTCT	CATGGGACCC	AGAGGTCTTC	CTGGTTCCCC	TGGAAATATC	1140
45	GGCCCCGCTG	GAAAAGAAGG	TCCTGTCGGC	CTCCCTGGCA	TCGACGGCAG	GCCTGGCCCA	1200
	ATTGGCCCAG	CTGGAGCAAG	AGGAGAGCCT	GGCAACATTG	GATTCCCTGG	ACCCAAAGGC	1260
50	CCCACTGGTG	ATCCTGGCAA	AAACGGTGAT	AAAGGTCATG	CTGGTCTTGC	TGGTGCTCGG	1320
7	GGTGCTCCAG	GTCCTGATGG	AAACAATGGT	GCTCAGGGAC	CTCCTGGACC	ACAGGGTGTT	1380

CAAGGTGGAA	AAGGTGAACA	GGGTCCCGCT	GGTCCTCCAG	GCTTCCAGGG	TCTGCCTGGC	1440
CCCTCAGGTC	CCGCTGGTGA	AGTTGGCAAA	CCAGGAGAĄA	GGGGTCTCCA	TGGTGAGTTT	1500
GGTCTCCCTG	GTCCTGCTGG	TCCAAGAGGG	GAACGCGGTC	CCCCAGGTGA	GAGTGGTGCT	1560
GCCGGTCCTA	CTGGTCCTAT	TGGAAGCCGA	GGTCCTTCTG	GACCCCCAGG	GCCTGATGGA	1620
AACAAGGGTG	AACCTGGTGT	GGTTGGTGCT	GTGGGCACTG	CTGGTCCATC	TGGTCCTAGT	1680
GGACTCCCAG	GAGAGAGGGG	TGCTGCTGGC	ATACCTGGAG	GCAAGGGAGA	AAAGGGTGAA	1740
CCTGGTCTCA	GAGGTGAAAT	TGGTAACCCT	GGCAGAGATG	GTGCTCGTGG	TGCTCATGGT	1800
GCTGTAGGTG	CCCCTGGTCC	TGCTGGAGCC	ACAGGTGACC	GGGCGAAGC	TGGGGCTGCT	1860
GGTCCTGCTG	GTCCTGCTGG	TCCTCGGGGA	AGCCCTGGTG	AACGTGGCGA	GGTCGGTCCT	1920
GCTGGCCCCA	ACGGATTTGC	TGGTCCGGCT	GGTGCTGCTG	GTCAACCGGG	TGCTAAAGGA	1980
GAAAGAGGAG	CCAAAGGGCC	TAAGGGTGAA	AACGGTGTTG	TTGGTCCCAC	AGGCCCCGTT	2040
GGAGCTGCTG	GCCCAGCTGG	TCCAAATGGT	CCCCCGGTC	CTGCTGGAAG	TCGTGGTGAT	2100
GGAGGCCCCC	CTGGTATGAC	TGGTTTCCCT	GGTGCTGCTG	GACGGACTGG	TCCCCCAGGA	2160
CCCTCTGGTA	TTTCTGGCCC	TCCTGGTCCC	CCTGGTCCTG	CTGGGAAAGA	AGGGCTTCGT	2220
GGTCCTCGTG	GTGACCAAGG	TCCAGTTGGC	CGAACTGGAG	AAGTAGGTGC	AGTTGGTCCC	2280
CCTGGCTTCG	CTGGTGAGAA	GGGTCCCTCT	GGAGAGGCTG	GTACTGCTGG	ACCTCCTGGC	2340
ACTCCAGGTC	CTCAGGGTCT	TCTTGGTGCT	CCTGGTATTC	TGGGTCTCCC	TGGCTCGAGA	2400

GGTGAACGTG	GTCTACCTGG	TGTTGCTGGT	GCTGTGGGTG	AACCTGGTCC	TCTTGGCATT	2460
GCCGGCCCTC	CTGGGGCCCG	TGGTCCTCCT	GGTGCTGTGG	GTAGTCCTGG	AGTCAACGGT	2520
GCTCCTGGTG	AAGCTGGTCG	TGATGGCAAC	CCTGGGAACG	ATGGTCCCCC	AGGTCGCGAT	2580
GGTCAACCCG	GACACAAGGG	AGAGCGCGGT	TACCCTGGCA	ATATTGGTCC	CGTTGGTGCT	2640
GCAGGTGCAC	CTGGTCCTCA	TGGCCCCGTG	GGTCCTGCTG	GCAAACATGG	AAACCGTGGT	2700
GAAACTGGTC	CTTCTGGTCC	TGTTGGTCCT	GCTGGTGCTG	TTGGCCCAAG	AGGTCCTAGT	2760
GGCCCACAAG	GCATTCGTGG	CGATAAGGGA	GAGCCCGGTG	AAAAGGGGCC	CAGAGGTCTT	2820
CCTGGCTTAA	AGGGACACAA	TGGATTGCAA	GGTCTGCCTG	GTATCGCTGG	TCACCATGGT	2880
GATCAAGGTG	CTCCTGGCTC	CGTGGGTCCT	GCTGGTCCTA	GGGGCCCTGC	TGGTCCTTCT	2940
GGCCCTGCTG	GAAAAGATGG	TCGCACTGGA	CATCCTGGTA	CGGTTGGACC	TGCTGGCATT	3000
CGAGGCCCTC	AGGGTCACCA	AGGCCCTGCT	GCCCCCTG	GTCCCCCTGG	CCCTCCTGGA	3060
CCTCCAGGTG	TAAGCGGTGG	TGGTTATGAC	TTTGGTTACG	ATGGAGACTT	CTACAGGGCT	3120

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1040 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

55

10

5

10

::

15

20

25

30

35

45

25	(xi)	SEQ	JENCI	E DES	SCRII	PTIO	N: S	EQ II	ои о	:30:						
	Gln 1	Туг	Asp	Gly	Lys 5	Gly	Val	Gly	Leu	Gly 10	Pro	Gly	Pro	Met	Gly 15	Leu
	Met	Gly	Pro	Arg 20	Gly	Pro	Pro	Gly	Ala 25	Ala	Gly	Ala	Pro	Gly 30	Pro	Gln
·15	Gly	Phe	Gln 35	Gly	Pro	Ala	Gly	Glu 40	Pro	Gly	Glu	Pro	Gly 45	Gln	Thr	Gly
20	Pro	Ala 50	Gly	Ala	Arg	Gly	Pro 55	Ala	Gly	Pro	Pro	Gly 60	Lys	Ala	Gly	Glu
25	Asp 65	Gly	His	Pro	Gly	Lys 70	Pro	Gly	Arg	Pro	Gly 75	Glu	Arg	Gly	Val	Val 80
30	Gly	Pro	Gln	Gly	Ala 85	Arg	Gly	Phe	Pro	Gly 90	Thr	Pro	Gly	Leu	Pro 95	Gly
35	Phe	Lys	Gly	Ile 100	Arg	Gly	His	Asn	Gly 105	Leu	Asp	Gly	Leu	Lys 110	Gly	Gln
40	Pro	Gly	Ala 115	Pro	Gly	Val	Lys	Gly 120	Glu	Pro	Gly	Ala	Pro 125	Gly	Glu	Asn
45	Gly	Thr	Pro	Gly	Gln	Thr	Gly 135	Ala	Arg	Gly	Leu	Pro 140	Gly	Glu	Arg	Gly
50	Arg	Val	Gly	Ala	Pro	Gly 150	Pro	Ala	Gly	Ala	Arg 155	Gly	Ser	Asp	Gly	Ser 160

	Val	Gly	Pro	Val	Gly	Pro	Ala	Gly	Pro	Ile	Gly	Ser	Ala	Gly	Pro	Pro
5					165					170					175	
P	Gly	Phe	Pro	Gly	Ala	Pro	Gly	Pro	Lys	Gly	Glu	Ile	Gly	Ala	Val	Gly
10	-			180			-		185	•			•	190		3
3.	Nan	Δla	Glv	Pro	Δla	Glv	Pro	פומ	Gly	Pro	Ara	G) v	Gl.v	val	01	Leu
	AU.	7124	195		7124	O.	110	200	Gry	710	AL 9	GLY	205	Val	GIY	Leu
15																
	Pro		Leu	Ser	Gly	Pro		Gly	Pro	Pro	Gly	Asn	Pro	Gly	Ala	Asn
20		210					215					220				
	Gly	Leu	Thr	Gly	Ala	Lys	Gly	Ala	Ala	Gly	Leu	Pro	Gly	Val	Ala	Gly
	225					230					235					240
25																
	Ala	Pro	Gly	Leu		Gly	Pro	Arg	Gly	Ile	Pro	Gly	Pro	Val	Gly	Ala
					245					250					255	
30	7. 1 -	Gly	פות	Th.~	Gl _v	21-	2	a 1		17. 1	a 1	Glu	D	01		
"	AIG	GIY	AIG	260	GIY	AIA	AIG	GIŞ	265	vai	GIÅ	Giu	Pro	270	Pro	Ala
														_,,		
35	Gly	Ser	Lys	Gly	Glu	Ser	Gly	Asn	Lys	Gly	Glu	Pro	Gly	Ser	Ala	Gly
7.			275					280					285			-
40 [.]	Pro	Gln	Gly	Pro	Pro	Gly	Pro	Ser	Gly	Glu	Glu	Gly	Lys	Arg	Gly	Pro
		290					295					300				
45	Asn	Gly	Glu	Ala	Gly	Ser	Ala	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Leu	Arq
	305				-	310		•			315			•		320
•																
50	Gly	Ser	Pro	Gly	Ser	Arg	Gly	Leu	Pro	Gly	Ala	Asp	Gly	Arg	Ala	Gly
<i>i</i> .					325					330					335	

	Val	Met	Gly	Pro	Pro	Gly	Ser	Arg	Gly	Ala	Ser	Gly	Pro	Ala	Gly	Val
5				340					345					350		
			_	_		_					21	93		~ 1	_	
	Arg	GIÀ	Pro	Asn	GIÀ	Asp	ATA		Arg	Pro	GIÅ	GIU		GIÀ	Leu	Met
10			355					360					365			
	Gly	Pro	Arg	Gly	Leu	Pro	Gly	Ser	Pro	Gly	Asn	Ile	Gly	Pro	Ala	Gly
15	•	370		_			375					380				•
:	Lys	Glu	Gly	Pro	Val	Gly	Leu	Pro	Gly	Ile	Asp	Gly	Arg	Pro	Gly	Pro
20	385					390					395					400
	Ile	Gly	Pro	Ala	Gly	Ala	Arg	Gly	Glu	Pro	Gly	Asn	Ile	Gly	Phe	Pro
25					405					410					415	
				_						_				_		
	Gly	Pro	Lys	-	Pro	Thr	Gly	Asp		Gly	Lys	Asn	Gly	_	Lys	Gly
				420					425					430		
30	'		41	•		a 1	• • • •		61	.1.		0 3	D	3	a1	•
	His	Ala	Gly	Leu	Ala	GIĄ	Ala	_	GIÅ	Ala	Pro	GTÅ	445	Авр	GIÀ	Asn
			435					440					443			
35	7 cn	Gly	Ala	Gln	Glv	Pro	Pro	Glv	Dro	Gl n	ردا م	Val	Gln	Glv	Glv	Tara
	ASII	450	ALU	0211	o _x		455	01		· · · ·	07	460		51	O ₂	_,,
<i>∵</i>		130														
40	Gly	Glu	Gln	Gly	Pro	Ala	Gly	Pro	Pro	Gly	Phe	Gln	Gly	Leu	Pro	Gly
	465			·		470	-				475					480
45	Pro	Ser	Gly	Pro	Ala	Gly	Glu	Val	Gly	Lys	Pro	Gly	Glu	Arg	Gly	Leu
•					485					490					495	
50	His	Gly	Glu	Phe	Gly	Leu	Pro	Gly	Pro	Ala	Gly	Pro	Arg	Gly	Glu	Arg
				500					505					510		

121

55

;

3 .	Gly	Pro	Pro	Gly	Glu	Ser	Gly	Ala	Ala	Gly	Pro	Thr	Gly	Pro	Ile	Gly
5			515					520					525			
•																
	Ser	Arg	Gly	Pro	Ser	Gly	Pro	Pro	Gly	Pro	Asp	Gly	Asn	Lys	Gly	Glu
10		530					535					540				
	Pro	Gly	Val	Val	Gly	Ala	Val	Gly	Thr	Ala	Gly	Pro	Ser	Gly	Pro	Ser
15	545					550					555					560
	Gly	Leu	Pro	Gly	Glu	Arg	Gly	Ala	Ala	Gly	Ile	Pro	Gly	Gly	Lys	Gly
20					565					570					575	
20																
•	Glu	Lys	Gly	Glu	Pro	Gly	Leu	Arg	Gly	Glu	Ile	Gly	Asn	Pro	Gly	Arg
N.				580					585					590		
25																
<i>:</i>	Asp	Gly	Ala	Arg	Gly	Ala	His	Gly	Ala	Val	Gly	Ala	Pro	Gly	Pro	Ala
r			595					600					605			
30											•					
	Gly		Thr	Gly	Asp	Arg		Glu	Ala	Gly	Ala	Ala	Gly	Pro	Ala	Gly
		610					615					620				
35				_	_		_	_								
•		Ala	GIY	Pro	Arg		Ser	Pro	Gly	Glu		GTÅ	Glu	Val	Gly	
	625					630					635					640
40	•••	61	D	>	61	DL.		-1			~1					_
	Ala	GIY	PIO	ASII		Pne	Ala	GIĀ	PIO		GIĀ	ATS	Ala	GIY	Gln	Pro
					645					650					655	
45	G) v	בות	Lva	G] v	Glu	7~~	Clv.	21-	T 120	C1	Dro	T	61	~1	B. a.m.	~1
45	Gry	AIG	цуз	660	Giu	λιy	Grå	ATG	665	GIY	PIO	nya	GIA		Asn	GIŸ
				000					003					670		
•	۷a۱	Val	Glv	Pro	Thr	G) v	Dro	۷»۱	G) v	11 =	Δls	G1v	Dro	בוג	Gly	Dro
50	742	744	675		••••	G.L.Y	-10	680	GLY	n.a	~TQ	GLY	685	vta	GIY	FIO
			973					990					000			

5	Asn	Gly 690	Pro	Pro	Gly	Pro	Ala 695	Gly	Ser	Arg	Gly	Asp 700	Gly	Gly	Pro	Pro
4																
10	Gly 705	Met	Thr	Gly	Phe	Pro 710	Gly	Ala	Ala	Gly	Arg 715	Thr	Gly	Pro	Pro	Gly 720
5																
15	Pro	Ser	Gly	Ile	Ser 725	Gly	Pro	Pro	Gly	Pro 730	Pro	Gly	Pro	Ala	Gly 735	Lys
	Glu	Gly	Leu		Gly	Pro	Arg	Gly		Gln	Gly	Pro	Val	_	Arg	Thr
20				740					745					750		
·	Gly	Glu	Val 755	Gly	Ala	Val	Gly	Pro 760	Pro	Gly	Phe	Ala	Gly 765	Glu	Lys	Gly
25																
•1	Pro		Gly	Glu	Ala	Gly		Ala	Gly	Pro	Pro		Thr	Pro	Gly	Pro
30		770					775					780				
:	Gln 785	Gly	Leu	Leu	Gly	Ala 790	Pro	Gly	Ile	Leu	Gly 795	Leu	Pro	Gly	Ser	Arg
35	,,,,					,,,,										
	Gly	Glu	Arg	Gly	Leu 805	Pro	Gly	Val	Ala	Gly 810	Ala	Val	Gly	Glu	Pro 815	Gly
40	Pro	Leu	Gly		Ala	Gly	Pro	Pro	_	Ala	Arg	Gly	Pro		Gly	Ala
				820					825					830		
45	Val	Gly	Ser 835	Pro	Gly	Val	Asn	Gly 840	Ala	Pro	Gly	Glu	Ala 845	Gly	Arg	Asp
50	Gly		Pro	Gly	Asn	Asp		Pro	Pro	Gly	Arg	-	Gly	Gln	Pro	Gly
••		850					855					860				

.	His	Lys	Gly	Glu	Arg	Gly	Tyr	Pro	Gly	Asn	Ile	Gly	Pro	Val	Gly	Ala
5	865					870					875					880
.	Ala	Gly	Ala	Pro	Gly	Pro	His	Gly	Pro	Val	Gly	Pro	Ala	Gly	Lys	His
, 10					885					890					895	
	Gly	Asn	Arg	Gly	Glu	Thr	Gly	Pro	Ser	Gly	Pro	Val	Gly	Pro	Ala	Gly
15				900					905					910		
,	Ala	Val	Gly	Pro	Arg	Gly	Pro	Ser	Gly	Pro	Gln	Gly	Ile	Arg	Gly	qaA
20			915					920					925			
	Lys	Gly	Glu	Pro	Gly	Glu	Lys	Gly	Pro	Arg	Gly	Leu	Pro	Gly	Leu	Lys
		930					935					940				
25	G 3	77.1 -	B	a 1	T	71 -	03	•	D	63	71 -		a 1	***		
	945	HIS	Asn	Gly	ren	950	GIY	Leu	Pro	GIŸ	955	Ala	GIĀ	HIS	HIS	9e0 GTÅ
30	,,,										,,,,					300
	Asp	Gln	Gly	Ala	Pro	Gly	Ser	Val	Gly	Pro	Ala	Gly	Pro	Arg	Gly	Pro
					965					970					975	
35	37	a 1	D	~~~	01	Dana		61	•	•	63		5 1	~ 3		_
ń	AIA	GIY	PIO	Ser 980	GIY	PIO	Ala	GIÀ	985 286	Asp	GIY	Arg	Thr	990	HIS	Pro
40	61	mb	1701	61. .	D	22.	G7.e.	-1 -	3	63	D	63 -	a 1		~1	~3
•	GIÀ	Int	995	Gly	PIO	ALA	GIY	1000		GIÀ	PIO	GIN	1005		GIN	GTĂ
45	Pro	Ala	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Val
		1010)				1019	5				1020)			,
50	Ser	Gly	Gly	Gly	Tyr	Asp	Phe	Gly	Tyr	Asp	Gly	Asp	Phe	Tyr	Arg	Ala
	1029	5				1030)				1039	5				1040

(2) INFORMATION FOR SEQ ID NO:31:

5

10

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3120 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

15

(ii) MOLECULE TYPE: cDNA

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

CAGTACGACG GTAAAGGCGT AGGCCTGGGT CCGGGTCCGA TGGGCCTGAT GGGTCCACGT 60 25 GGCCCACCGG GTGCAGCAGG TGCGCCGGGT CCGCAGGGCT TCCAAGGTCC GGCGGGTGAA CCGGGCGAAC CGGGTCAGAC GGGTCCGGCG GGTGCTCGCG GTCCGGCTGG CCCACCGGGC 180 30 AAAGCTGGCG AAGACGGTCA CCCGGGTAAG CCAGGCCGCC CGGGCGAACG TGGCGTCGTG 240 35 GGTCCGCAAG GTGCGCGTGG TTTCCCGGGC ACGCCGGGTC TGCCGGGTTT CAAAGGCATT 300 CGTGGTCACA ACGGTCTGGA CGGTCTGAAA GGCCAACCGG GTGCTCCGGG CGTCAAAGGC 360 40 GAACCGGGTG CCCCAGGCGA AAACGGTACG CCGGGCCAGA CTGGTGCGCG TGGTCTGCCG 420 GGTGAACGCG GCCGTGTTGG CGCTCCGGGT CCGGCTGGCG CGCGTGGCAG CGATGGCTCC 480 45 GTCGGTCCGG TTGGCCCTGC GGGTCCGATT GGTTCCGCTG GCCCTCCGGG TTTCCCGGGT 540 GCGCCGGGTC CGAAGGGTGA GATCGGCGCG GTTGGCAACG CAGGCCCGGC TGGTCCAGCC 600

GGCCCTCGTG GCGAAGTCGG TCTGCCGGGT CTGAGCGGTC CGGTAGGCCC ACCGGGTAAC

55

 $f \rightarrow$

՝

720	CGTTGCCGGT	GCCTGCCGGG	GGTGCGGCTG	GGGTGCAAAA	ACGGCCTGAC	CCGGGCGCAA	
780	CGGTGCAACT	TAGGCGCAGC	ссеетссее	GCGCGGTATT	TGCCGGGTCC	GCCCGGGCC	
840	AAGCGGTAAC	CTAAAGGCGA	CCGGCGGGTT	CGAACCGGGT	GCCTGGTTGG	GGTGCCCGTG	
900	CGAAGAAGGT	GTCCGAGCGG	GGTCCGCCGG	GGGCCCGCAG	CGGGTTCCGC	AAAGGTGAGC	
960	GGGTCTGCGT	CGGGTCCGCC	GCAGGCCCTC	GGCTGGTTCC	CGAACGGCGA	AAACGTGGTC	
1020	GATGGGTCCG	GTGCGGGCGT	GCGGACGGCC	CCTGCCGGGC	GTAGCCGTGG	GGCAGCCCGG	
1080	CGACGCGGGC	GTCCGAATGG	GGTGTCCGTG	TGGTCCGGCT	GTGGTGCCTC	CCGGGTTCCC	
1140	GGGTAACATT	CGGGTAGCCC	CGTGGCCTGC	GATGGGTCCG	AACCGGGCCT	CGTCCGGGTG	
1200	TCCGGGTCCG	TTGATGGTCG	CTGCCGGGTA	TCCGGTAGGT	GTAAGGAGGG	GGTCCGGCGG	
1260	TCCGAAGGGT	GTTTTCCGGG	GGTAACATCG	TGGCGAGCCG	CGGGCGCTCG	ATCGGCCCTG	
1320	AGGTGCCCGT	CAGGTCTGGC	AAAGGCCATG	GAACGGTGAT	ACCCGGGCAA	CCGACGGGCG	
1380	GCAGGGCGTA	CGCCGGGTCC	GCGCAGGGTC	TAACAATGGT	GTCCGGATGG	GGTGCACCGG	
1440	TCTGCCGGGT	GCTTCCAGGG	GGCCCACCGG	GGGTCCGGCA	AAGGTGAACA	CAGGGTGGCA	
1500	TGGCGAGTTT	GTGGCCTCCA	CCGGGCGAAC	AGTGGGCAAA	CGGCTGGTGA	CCGAGCGGCC	
1560	ATCCGGCGCG	CTCCGGGCGA	GAGCGCGGCC	TCCGCGTGGT	GTCCGGCCGG	GGCCTGCCGG	
1620	TCCGGACGGC	GCCCACCGGG	GGTCCGAGCG	TGGTTCCCGT	CCGGCCCGAT	GCAGGTCCGA	
1680	TGGTCCGAGC	сседсссетс	GTTGGTACCG	TGTTGGTGCT	AGCCGGGTGT	AACAAAGGCG	

· 5

·₁₅

.;

GGTCTGCCGG	GCGAACGCGG	TGCCGCTGGT	ATTCCGGGCG	GCAAAGGTGA	AAAAGGTGAA	1740
CCGGGTCTGC	GCGGTGAGAT	TGGCAACCCG	GGCCGTGAÇG	GTGCTCGCGG	TGCACACGGC	1800
GCGGTTGGCG	CACCGGGTCC	GGCAGGCGCG	ACTGGTGATC	GTGGCGAAGC	TGGTGCAGCG	1860
GGTCCGGCGG	GTCCGGCCGG	CCCTCGCGGT	TCCCCGGGCG	AACGCGGCGA	AGTCGGCCCG	1920
GCTGGCCCGA	ATGGCTTTGC	TGGCCCAGCG	GGCGCTGCGG	GCCAACCGGG	TGCGAAAGGT	1980
GAGCGCGGTG	CCAAAGGCCC	GAAAGGTGAA	AATGGTGTAG	TTGGTCCGAC	GGGTCCGGTT	2040
GGTGCGGCTG	GTCCGGCTGG	CCCGAATGGT	CCGCCGGGTC	CGGCAGGCAG	CCGTGGCGAT	2100
GGTGGCCCAC	CGGGCATGAC	CGGTTTCCCT	GGCGCGGCCG	GTCGCACCGG	CCCGCCGGGT	2160
CCGTCTGGCA	TTTCTGGCCC	ACCGGGTCCG	CCGGGTCCGG	CGGGCAAAGA	AGGTCTGCGT	2220
GGCCCACGCG	GCGACCAGGG	TCCGGTGGGC	CGTACCGGCG	AAGTCGGTGC	TGTTGGCCCT	2280
CCGGGCTTTG	CGGGTGAGAA	AGGTCCGAGC	GGTGAAGCTG	GCACCGCAGG	CCCGCCGGGT	2340
ACGCCGGGTC	CGCAAGGTCT	GCTGGGTGCT	CCGGGTATCC	TGGGCCTGCC	GGGCTCCCGT	2400
GGCGAACGCG	GTCTGCCGGG	CGTTGCAGGC	GCTGTAGGCG	AACCGGGTCC	GCTGGGTATC	2460
GCGGGTCCGC	CGGGTGCGCG	TGGTCCGCCG	GGTGCCGTGG	GCTCTCCGGG	TGTTAACGGC	2520
GCCCCTGGTG	AAGCGGGCCG	CGACGGCAAT	CCGGGCAACG	ATGGTCCGCC	GGGTCGTGAT	2580
GGTCAGCCGG	GTCACAAAGG	TGAGCGTGGC	TACCCGGGTA	ACATCGGTCC	GGTTGGTGCG	2640
GCCGGCGCTC	CGGGTCCGCA	CGGTCCGGTA	GGCCCAGCCG	GCAAACACGG	TAACCGTGGT	2700

.:

::

	.:	GAAACGGGTC CGTCCGGTCC GGTAGGTCCG GCGGGTGCTG TTGGTCCACG CGGCCCGTCC	2760
GATCAGGGTG CTCCGGGTTC CGTTGGTCCG GCCGGTCCGC GTGGCCCGGC TGGTCCGTCT GGTCCGGCCG GTAAAGACGG CCGTACGGC CACCCGGGTA CGGTGGGTCC GGCCGGCATT CGCGGTCCGC AAGGTCACCA GGGTCCGGCG GGTCCGCCGG TCCGCCGGGT CCGCCGGGTG TTAGCGGTGG CGGTTATGAT TTTGGTTATG ACGGTGATTT CTATCGTGCG (2) INFORMATION FOR SEQ ID NO:32: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1040 amino acids (B) TYPE: amino acid (C) STRANDEDNESS: single (D) TOPOLOGY: unknown (ii) MOLECULE TYPE: peptide (xi) SEQUENCE DESCRIPTION: SEQ ID NO:32: Gln Tyr Asp Gly Lys Gly Val Gly Leu Gly Pro Gly Pro Met Gly Leu	5	GGCCCGCAGG GTATTCGCGG TGACAAAGGC GAACCGGGCG AAAAAGGTCC GCGTGGTCTG	2820
GGTCCGGCCG GTAAAGACGG CCGTACGGGC CACCCGGGTA CGGTGGGTCC GGCCGGCATT CGCGGGTCCGC AAGGTCACCA GGGTCCGGCG GGTCCGCCGGG TCCGCCGGGT CCGCCGGGTG TTAGCGGTGG CGGTTATGAT TTTGGTTATG ACGGTGATTT CTATCGTGCG (2) INFORMATION FOR SEQ ID NO:32: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1040 amino acids (B) TYPE: amino acid (C) STRANDEDNESS: single (D) TOPOLOGY: unknown (ii) MOLECULE TYPE: peptide (xi) SEQUENCE DESCRIPTION: SEQ ID NO:32: Gln Tyr Asp Gly Lys Gly Val Gly Leu Gly Pro Gly Pro Met Gly Leu	10	CCGGGCCTTA AGGGCCACAA CGGTCTGCAA GGTCTGCCGG GTATCGCGGG TCACCACGGT	2880
CGCGGTCCGC AAGGTCACCA GGGTCCGCGG GTCCGCCGGG TCCGCCGGGT CCGCCGGGTG TTAGCGGTGG CGGTTATGAT TTTGGTTATG ACGGTGATTT CTATCGTGCG (2) INFORMATION FOR SEQ ID NO:32: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1040 amino acids (B) TYPE: amino acid (C) STRANDEDNESS: single (D) TOPOLOGY: unknown (ii) MOLECULE TYPE: peptide (xi) SEQUENCE DESCRIPTION: SEQ ID NO:32: Gln Tyr Asp Gly Lys Gly Val Gly Leu Gly Pro Gly Pro Met Gly Leu	ť.	GATCAGGGTG CTCCGGGTTC CGTTGGTCCG GCCGGTCCGC GTGGCCCGGC TGGTCCGTCT	2940
CCGCCGGTC TAGCCGTGG CGGTTATGAT TTTGGTTATG ACGGTGATTT CTATCGTGCG CCGCCGGGTG TTAGCCGTGG CGGTTATGAT TTTGGTTATG ACGGTGATTT CTATCGTGCG (2) INFORMATION FOR SEQ ID NO:32: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1040 amino acids (B) TYPE: amino acid (C) STRANDEDNESS: single (D) TOPOLOGY: unknown (ii) MOLECULE TYPE: peptide (xi) SEQUENCE DESCRIPTION: SEQ ID NO:32: Gln Tyr Asp Gly Lys Gly Val Gly Leu Gly Pro Gly Pro Met Gly Leu	15	GGTCCGGCCG GTAAAGACGG CCGTACGGGC CACCCGGGTA CGGTGGGTCC GGCCGGCATT	3000
CCGCCGGGTG TTAGCGGTGG CGGTTATGAT TTTGGTTATG ACGGTGATTT CTATCGTGCG (2) INFORMATION FOR SEQ ID NO:32: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1040 amino acids (B) TYPE: amino acid (C) STRANDEDNESS: single (D) TOPOLOGY: unknown (ii) MOLECULE TYPE: peptide (xi) SEQUENCE DESCRIPTION: SEQ ID NO:32: Gln Tyr Asp Gly Lys Gly Val Gly Leu Gly Pro Gly Pro Met Gly Leu		CGCGGTCCGC AAGGTCACCA GGGTCCGGCG GGTCCGCCGGG TCCGCCGGGT	3060
(2) INFORMATION FOR SEQ ID NO:32: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1040 amino acids (B) TYPE: amino acid (C) STRANDEDNESS: single (D) TOPOLOGY: unknown (ii) MOLECULE TYPE: peptide (xi) SEQUENCE DESCRIPTION: SEQ ID NO:32: Gln Tyr Asp Gly Lys Gly Val Gly Leu Gly Pro Gly Pro Met Gly Leu		CCGCCGGGTG TTAGCGGTGG CGGTTATGAT TTTGGTTATG ACGGTGATTT CTATCGTGCG	3120
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 1040 amino acids (B) TYPE: amino acid (C) STRANDEDNESS: single (D) TOPOLOGY: unknown (ii) MOLECULE TYPE: peptide (xi) SEQUENCE DESCRIPTION: SEQ ID NO:32: Gln Tyr Asp Gly Lys Gly Val Gly Leu Gly Pro Gly Pro Met Gly Leu		(2) INFORMATION FOR SEQ ID NO:32:	
(ii) MOLECULE TYPE: peptide (xi) SEQUENCE DESCRIPTION: SEQ ID NO:32: Gln Tyr Asp Gly Lys Gly Val Gly Leu Gly Pro Gly Pro Met Gly Leu	30	(A) LENGTH: 1040 amino acids(B) TYPE: amino acid(C) STRANDEDNESS: single	
45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:32: Gln Tyr Asp Gly Lys Gly Val Gly Leu Gly Pro Gly Pro Met Gly Leu			
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32: Gln Tyr Asp Gly Lys Gly Val Gly Leu Gly Pro Gly Pro Met Gly Leu		(ii) MOLECULE TYPE: peptide	
Gln Tyr Asp Gly Lys Gly Val Gly Leu Gly Pro Gly Pro Met Gly Leu	45	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:	
1 5 10 15	50	Gln Tyr Asp Gly Lys Gly Val Gly Leu Gly Pro Gly Pro Met Gly Leu 1 5 10 15	

.:	Met	Gly	Pro	Arg	Gly	Pro	Pro	Gly	Ala	Ala	Gly	Ala	Pro	Gly	Pro	Gln
5				20					25					30		
i. 10	Gly	Phe	Gln 35	Gly	Pro	Ala	Gly	Glu 40	Pro	Gly	Glu	Pro	Gly 45	Gln	Thr	Gly
15	Pro	Ala 50	Gly	Ala	Arg	Gly	Pro 55	Ala	Gly	Pro	Pro	Gly 60	Lys	Ala	Gly	Glu
	Asp 65	Gly	His	Pro	Gly	Lys 70	Pro	Gly	Arg	Pro	Gly 75	Glu	Arg	Gly	Val	Val
20	Gly	Pro	Gln	Gly	Ala 85	Arg	Gly	Phe	Pro	Gly 90	Thr	Pro	Gly	Leu	Pro 95	Gly
25	Phe	Lys	Gly	Ile	Arg	Gly	His	Asn	Gly 105	Leu	Asp	Gly	Leu	Lys 110	Gly	Gln
30	Pro	Gly	Ala 115	Pro	Gly	Val	Lys	Gly 120	Glu	Pro	Gly	Ala	Pro 125	Gly	Glu	Asn
35	Gly	Thr		Gly	Gln	Thr	Gly	Ala	Arg	Gly	Leu	Pro	Gly	Glu	Arg	Gly
40	Arg		Glý	Ala	Pro	Gly 150	Pro	Ala	Gly	Ala	Arg 155	Gly	Ser	Asp	Gly	Ser
45	Val	Gly	Pro		Gly 165		Ala	Gly	Pro	Ile 170	Gly	Ser	Ala	Gly	Pro 175	Pro
50	Gly	Phe	Pro	Gly 180		Pro	Gly	Pro	Lys 185		Glu	Ile	Gly	Ala 190	Val	Gly

ţ.

	Asn	Ala	Gly	Pro	Ala	Gly	Pro	Ala	Gly	Pro	Arg	Gly	Glu	Val	Gly	Leu
5			195					200					205			
			_			_	•		_		~ 3		_			
	Pro		Leu	Ser	Gly	Pro		Gly	Pro	Pro	GIA		Pro	GΙΆ	Ala	Asn
10		210					215					220				
		•	m la sa	~ 1		T	a1	210		~1	T 011	D	71. .	V-1	••-	
	-	ren	inr	GIA	Ala	230	GIY	ALA	MIG	GIY	235	PLO	GTA	val	Ala	-
15	225					230					233					240
	Ala	Pro	Glv	Leu	Pro	Glv	Pro	Ara	Glv	Ile	Pro	Glv	Pro	Val	Glv	Ala
			,	_ •	245				•	250		•			255	
20									•							
	Ala	Gly	Ala	Thr	Gly	Ala	Arg	Gly	Leu	Val	Gly	Glu	Pro	Gly	Pro	Ala
		•		260					265					270		
0.5																
25	Gly	Ser	Lys	Gly	Glu	Ser	Gly	Asn	Lys	Gly	Glu	Pro	Gly	Ser	Ala	Gly
•			275					280					285			
30	Pro	Gln	Gly	Pro	Pro	Gly	Pro	Ser	Gly	Glu	Glu	Gly	Lys	Arg	Gly	Pro
		290					295					300				
35	Asn	Gly	Glu	Ala	Gly	Ser	Ala	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Leu	Arg
	305					310					315					320
	=	_	_		_	_	b		_			_		_		
40	Gly	Ser	Pro	Gly		Arg	Gly	Leu	Pro	_	Ala	Asp	GTA	Arg		Gly
					325					330					335	
	17-1	Mat	alv	Pro	Pro	G1 v	Car	7~~	alv	פומ	Car	Glv	Dro	a la	Gly	Val
	val	Mec	GIY	340	PLO	GIY	261	Atg	345	ATG	361	GIY	210	350	GIY	val
45				240					343					,,,,		
	Ara	Glv	Pro	Asn	Glv	Asp	Ala	Glv	Ara	Pro	Glv	Glu	Pro	Glv	Leu	Met
	3	•	355		•	•		360	3				365	3		•
50																

·	Gly	Pro	Arg	Gly	Leu	Pro	Gly	Ser	Pro	Gly	Asn	Ile	Gly	Pro	Ala	Gly
5		370					375					380				
	Lys	Glu	Gly	Pro	Val		Leu	Pro	Gly	Ile		Gly	Arg	Pro	Gly	Pro
10	385					390					395					400
	Ile	Gly	Pro	Ala		Ala	Arg	Gly	Glu		Gly	Asn	Ile	Gly		Pro
15					405					410					415	
	Gly	Pro	Lys	Gly 420	Pro	Thr	Gly	Asp	Pro	Gly	Lys	Asn	Gly	Asp 430	Lys	Gly
20				420					425					430		
	His	Ala	Gly 435	Leu	Ala	Gly	Ala	Arg 440	Gly	Ala	Pro	Gly	Pro 445	Asp	Gly	Asn
25						_				_	_	_				
	Asn	G1y 450	Ala	Gln	Gly	Pro	Pro 455	Gly	Pro	Gln	Gly	Val 460	Gln	Gly	Gly	Lys
30	G] v	el u	Gln.	al v	Pro	חות	Glv.	Dwa	Dwa	<i>~</i> 1	Dho	61 m	G3	Leu	D	63
	465	GIU	GIII	GLY	PIO	470	GIY	PIO	PIO	GIY	475	GIII	GIŸ	reu	PIO	480
35	Pro	Ser	Gly	Pro	Ala	Gly	Glu	Val	Gly	Lys	Pro	Gly	Glu	Arg	Glv	Leu
			-		485	-			-	490		-		•	495	
. · 40	His	Gly	Glu	Phe	Gly	Leu	Pro	Gly	Pro	Ala	Gly	Pro	Arg	Gly	Glu	Arg
				500					505					510		
45	Gly	Pro	Pro	Gly	Glu	Ser	Gly	Ala	Ala	Gly	Pro	Thr	Gly	Pro	Ile	Gly
			515					520					525			
50	Ser		Gly	Pro	Ser	Gly		Pro	Gly	Pro	Asp		Asn	Lys	Gly	Glu
		530					535					540				

55·

	Pro	Gly	Val	Val	Gly	Ala	Val	Gly	Thr	Ala	Gly	Pro	Ser	Gly	Pro	Ser
5	545					550					555					560
	Gly	Leu	Pro	Glý	Glu	Arg	Gly	Ala	Ala	Gly	Ile	Pro	Gly	Gly	Lys	Gly
10					565		_			570			_	_	- 575	•
	Glu	T.vs	Glv	Glu	Pro	Glv	T.em	Δτα	Glv	Glu	Tle	Glv	Δgn	Dro	G1.	N
	014	_,_	,	580		1		•9	585	010		017		590	Gry	Arg
15				300					303					330		
	7.00	C1.,	71-	Arg	Clv.	71-	ni o	~1··	210	170 1	~1	210	77-0	~1	D	
	Азр	GIY		ALG.	GIY	AId	urs		AIA	Val	GIY	AIA		GIY	Pro	ATA
20			595					600					605			
	_				_			_	_	_	_	_	_			
	Gly		Thr	Gly	qaA	Arg		Glu	Ala	Gly	Ala		Gly	Pro	Ala	Gly
0.5		610					615					620				
25		•														
	Pro	Ala	Gly	Pro	Arg	Gly	Ser	Pro	Gly	Glu	Arg	Gly	Glu	Val	Gly	Pro
	625					630					635					640
30																
	Ala	Gly	Pro	Asn	Gly	Phe	Ala	Gly	Pro	Ala	Gly	Ala	Ala	Gly	Gln	Pro
;					645					650					655	
35 ⁷																
30	Gly	Ala	Lys	Gly	Glu	Arg	Gly	Ala	Lys	Gly	Pro	Lys	Gly	Glu	Asn	Gly
				660					665					670		
,																
40	Val	Val	Gly	Pro	Thr	Gly	Pro	Val	Gly	Ala	Ala	Gly	Pro	Ala	Gly	Pro
			675			-		680				•	685		•	
45	Asn	Glv	Pro	Pro	Glv	Pro	Ala	Glv	Ser	Ara	Glv	Asn	Glv	Glv	Pro	Pro
		690			1		695	,		5	,	700	023	O ₂		110
												, 50				
	GI v	Met	ሞኮ∽	alv	Dhe	Dro	G1··	71~	77-	C1	3	መት	01	Dece	D	~1
50		MEL	1111	Gly	File		GIÅ	wrg	WTS	σтλ		Int	GTÅ	PLO	LLO	
	705					710					715					720

. 5	Pro	Ser	Gly	Ile	Ser 725	Gly	Pro	Pro	Gly	Pro 730	Pro	Gly	Pro	Ala	Gly 735	Lys
: 10	Glu	Gly	Leu	Arg 740	Gly	Pro	Arg	Gly	Asp 745	Gln	Gly	Pro	Val	Gly 750	Arg	Thr
:	Gly	Glu	Val 755	Gly	Ala	Val	Gly	Pro 760	Pro	Gly	Phe	Ala	Gly 765	Glu	Lys	Gly
15	Pro			Glu	Ala	Gly			Gly	Pro	Pro	_		Pro	Gly	Pro
20	Gln	770 Gly	Leu	Leu	Gly	Ala	775 Pro	Gly	Ile	Leu	Gly	780 Leu	Pro	Gly	Ser	Arg
25	785 Gly	Glu	Arg	Gly	Leu	790 Pro	Gly	Val	Ala	Gly	795 Ala	Val	Gly	Glu	Pro	800 Gly
30	Pro	Leu	Gly	Ile	805 Ala	Glv	Pro	Pro	Glv	810 Ala	Arq	Glv	Pro	Pro	815 Gly	Ala
:· 35				820					825					830	-	
9 40	vai	GIY	835	PIO	GIY	vai	Asn	840	Ala	Pro	GIÀ	GIU	845	GIÀ	Arg	Asp
	Gly	850	Pro	Gly	Asn	qaA	Gly 855	Pro	Pro	Gly	Arg	Asp	Gly	Gln	Pro	Gly
45	His 865	Lys	Gly	Glu	Arg	Gly 870	Tyr	Pro	Gly	Asn	Ile 875	Gly	Pro	Val	Gly	Ala 880
50	Ala	Gly	Ala	Pro	Gly 885	Pro	His	Gly	Pro	Val 890	Gly	Pro	Ala	Gly	Lys 895	His

	Gly	Asn	Arg	Gly	Glu	Thr	Gly	Pro	Ser	Gly	Pro	Val	Gly	Pro	Ala	Gly
5				900					905					910		
	Ala	Val	Gly	Pro	Arg	Gly	Pro	Ser	Gly	Pro	Gln	Gly	Ile	Arg	Gly	Asp
10			915					920					925		·	_
	T va	Gly	Gl u	Dro	Gl.v	C1	T 110	~1	Dwa	2	C1	T 011	D	01	_	_
	Lys	930	GIU	PIO	GIY	Gru	935	GIY	PIO	Arg	GIY	940	PIO	GIÅ	Leu	Lys
15																
	Gly	His	Asn	Gly	Leu	Gln	Gly	Leu	Pro	Gly	Ile	Ala	Gly	His	His	Gly
	945					950					955					960
20	3	a 1-	61	21-	Dava	~ 1		••- •	-3				_	_		
	Asp	Gln	GIY	Ala	965	GIÀ	ser	Val	GIÀ	970	Ala	GIÀ	Pro	Arg	Gly 975	Pro
25										3,0					313	
	Ala	Gly	Pro	Ser	Gly	Pro	Ala	Gly	Lys	Asp	Gly	Arg	Thr	Gly	His	Pro
				980					985					990		
30			-		_											
	Gly	Thr	995	GIÀ	Pro	Ala	Gly	Ile 1000		Gly	Pro	Gln	Gly 1005		Gln	Gly
			,,,					1000	,				1003	•		
35	Pro	Ala	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Val
		1010)				1015	;				1020)			
		~ 3	01	a 1												
10	ser 102!	Gly	GIÅ	GIÅ	Tyr	Asp 1030		Gly	Tyr	Asp	Gly 1035		Phe	Tyr	Arg	
:	202.					1030					1035	,				1040
1 5																
	(2) INFO	RMATI	ON F	FOR S	EQ I	D NC):33:									
•		000-	·		. n											
50	(1)	SEQU (A)		CHA IGTH :												
				е: г			_									
				ANDE		-		.e								
55																

5	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: cDNA	
10	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:	
ę	GGAATTCATG CAGTATGATG GCAAAGGCGT CGGCCTCGGC CCGGGCCCAA TGGGCCTCAT	60
	GGGCCCGCGC GGCCCA	76
20	(2) INFORMATION FOR SEQ ID NO:34:	
	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 79 base pairs	
25	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: single	
<i>A</i>	(D) TOPOLOGY: linear	
30		
*	(ii) MOLECULE TYPE: cDNA	
<i>ti</i>	· ·	
35		
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:	
40	CCGGGCGCC CGGGTGCCC ACGTCGACCG CGGGGTCCGG GCGTTCCAAA GGTCCCGGGA	60
45	CGGCCAATTA TTCGAACCC	79
	(2) INFORMATION FOR SEQ ID NO:35:	
50	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 82 base pairs	
•	(B) TYPE: nucleic acid	
55	• •	

	(C) STRANDEDNESS: single	
5 ·	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: cDNA	
10	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:	
·15·	GGAATTCGCC GGTGAGCCGG GTGAACCGGG CCAAACGGGT CCGGCAGGTC CACGTGGTCC	60
	AGCGGGCCCG CCTGGCAAGG CG	82
20	(2) INFORMATION FOR SEQ ID NO:36:	
	(i) SEQUENCE CHARACTERISTICS:	•
25 ⁻	(A) LENGTH: 84 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: single	
30	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: cDNA	
35 [.]		
00		
.:		
40·	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:	
•	CCGGGCGGAC CGTTCCGCCC ACTTCTACCG GTGGGACCGT TTGGCCCGGC GGGCCACTCG	60
45 [.]	CACCGCATCA CATTATTCGA ACCC	84
	(2) INFORMATION FOR SEQ ID NO:37:	
50	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 240 base pairs	
	(A) DENGIN. 210 DESC PAILS	

(B) TYPE: nucleic acid

	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
10	(ii) MOLECULE TYPE: cDNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:	
: 15		
•	CAGTATGATG GCAAAGGCGT CGGCCTCGGC CCGGGCCCAA TGGGCCTCAT GGGCCCGCGC	60
	GGCCCACCGG GTGCAGCTGG CGCCCCAGGC CCGCAAGGTT TCCAGGGCCC TGCCGGTGAG	L20
20		
	CCGGGTGAAC CGGGCCAAAC GGGTCCGGCA GGTGCACGTG GTCCAGCGGG CCCGCCTGGC	180
25	AAGGCGGGTG AAGATGGCCA CCCTGGCAAA CCGGGCCGCC CGGGTGAGCG TGGCGTAGTG	240
	(2) INFORMATION FOR SEQ ID NO:38:	
30	(a)	
	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 80 amino acids	
35 [°]	(B) TYPE: amino acid	
	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: unknown	
	(b) 10102001. madiown	
40	(ii) MOLECULE TYPE: peptide	
	(11, the Doom I than population	
45		
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:	
50	Gln Tyr Asp Gly Lys Gly Val Gly Leu Gly Pro Gly Pro Met Gly Leu	
	1 5 10 15	
55		

5	Met	Gly	Pro	Arg 20	Gly	Pro	Pro	Gly	Ala 25	Ala	Gly	Ala	Pro	Gly 30	Pro	Gln	
v.	Gly	Phe	Gln	Gly	Pro	Ala	Gly	Glu	Pro	Gly	Glu	Pro	Gly	Gln	Thr	Gly	
10			35					40	•				45				
15	Pro	Ala 50	Gly	Ala	Arg	Gly	Pro 55	Ala	Gly	Pro	Pro	Gly 60	Lys	Ala	Gly	Glu	
	Asp	Gly	His	Pro	Gly	Lys	Pro	Gly	Arg	Pro	Gly	Glu	Arg	Gly	Val	Val	
20	65					70					75					80	
25	(2) INFO	RMAT:	ION 1	FOR S	SEQ :	ID NO	D:39	:									•
••	(i)			E CHI NGTH													
30		(B)) TY	PE: 1	nucle	eic a	acid										
ι <u>,'</u>				RANDI POLO			_	le									
35	(ii)	MOL	ECULI	E TY	PE: 0	odna											
40	(xi)	SEQ	UENCI	E DES	SCRI	PTIO	N: SI	II QE	O NO:	:39:							
	ATGGGGCT	cg c	rggc(CCAC	C GG(GCGAI	ACCG	GGT	CCGC	CAG (SCCC	AAAE	G T	CCGC	STGGC	2	60
45	GATAGCGG	GC T	CGCT	3GCC(C AC	CGGG(CGAA	CCGC	GTC	CGC (CAGG	CCG	AA AG	GTC	CGCG1	ŗ	120
50	GGCGATAG	CG G(GCTC	3CTG(3 CC	CACCO	GGC	GAAG	CCGG	GTC (CGCC#	/GGC(CC G/	LAA G(TCCG	3	180
	CGTGGCGA	ra G	CGGG	CTCG	C TG(3CCC1	ACCG	GGC	BAAC(CGG (STCC	GCAC	G C	CCGAI	\AGG1	5	240
55	CCGCGTGG	CG A	TAGC	3GGC	r cc	CGGG	CGAT	TCC	raa								276

	(2)	INFO	TAMS	ION 1	FOR S	SEQ :	ID N	0:40	:								
. 5																	
, f		(i)	SEQU	JENCI	E CHI	ARAC:	reri:	STIC	S:		•						
			(A)	LEI	NGTH	: 91	amiı	no a	cids								
10			(B)	TYI	PE: 8	amino	o ac	id									
						EDNES		•	le								
			(D)	TOI	POLO	3Y: ι	ınkno	own									
15		(ii)	MOLI	CULI	E TYI	PE: I	pept:	ide									
		(xi)	SEQU	JENCI	E DES	SCRII	PTIO	N: SI	EQ II	ON C	:40:						
20																	
·		Met	Gly	Leu	Ala	Glý	Pro	Pro	Gly	Glu	Pro	Gly	Pro	Pro	Gly	Pro	Lys
		1				5					10					15	
25		_												_			
		Gly	Pro	Arg		Asp	Ser	Gly	Leu		Gly	Pro	Pro	Gly		Pro	Gly
					20					25					30		
30		Pro	Pro	ตา v	Pro	I.vs	Glv	Pro	Ara	Glv	Agn	Ser	G] v	T.em	A1=	Gly	D×0
;		110		35		273	O.J		40	Q-y	wab	501	U.J	45	AL Q	GLY	FIO
35		Pro	Gly	Glu	Pro	Gly	Pro	Pro	Gly	Pro	Lys	Gly	Pro	Arg	Gly	Asp	Ser
j			50					55					60				
40		Gly	Leu	Ala	Gly	Pro	Pro	Gly	Glu	Pro	Gly	Pro	Pro	Gly	Pro	Lys	Gly
		65					70					75					80
							•										
45		Pro	Arg	Gly	Asp	Ser	Gly	Leu	Pro	Gly	Asp	Ser					
						85					90						

(2) INFORMATION FOR SEQ ID NO:41:

5	
	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 13 amino acids
	(B) TYPE: amino acid
10	(C) STRANDEDNESS: single
	(D) TOPOLOGY: unknown
37	
15	(ii) MOLECULE TYPE: peptide
.*	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:
20	
	Gly Pro Pro Gly Leu Ala Gly Pro Pro Gly Glu Ser Gly
	1 5 10
25	(2) INFORMATION FOR SEQ ID NO:42:
	,-,
	(i) SEQUENCE CHARACTERISTICS:
30	(A) LENGTH: 13 amino acids
•	(B) TYPE: amino acid
	(C) STRANDEDNESS: single
35	(D) TOPOLOGY: unknown
	(ii) MOLECULE TYPE: peptide
ਦ 40	
	(ix) FEATURE:
	(A) NAME/KEY: Modified-site
45 [:]	(B) LOCATION: 23
	(D) OTHER INFORMATION: /product= "4-hydroxyproline"
50	

di.	(ix) FEATURE:	
·· 5	(A) NAME/KEY: Modified-site	
	(B) LOCATION: 89	
	(D) OTHER INFORMATION: /product= "Xaa = 4-hydroxyproline"	
J		
10	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:	
	Gly Xaa Xaa Gly Leu Ala Gly Xaa Xaa Gly Glu Ser Gly	
15	1 5 10	
	(2) INFORMATION FOR SEQ ID NO:43:	
20		
•	(i) SEQUENCE CHARACTERISTICS:	•
**	(A) LENGTH: 660 base pairs	
25	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
,		
30	(ii) MOLECULE TYPE: cDNA	
s	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:	
35	ATGGGCCCGC CGGGTCTGGC GGGCCCTCCG GGTGAAAGCG GTCGTGAAGG CGCGCCGGGT	60
40	GCCGAAGGCA GCCCAGGCCG CGACGGTAGC CCGGGGGCCA AAGGGGATCG TGGTGAAACC	120
	GGCCCGGCGG GCCCCCGGG TGCACCGGGC GCGCCGGGTG CCCCAGGCCC GGTGGGCCCG	180
45	GCGGGCAAAA GCGGTGATCG TGGTGAGACC GGTCCGGCGG GCCCGGCCGG TCCGGTGGGC	240
50	CCAGCGGGCG CCCGTGGCCC GGCCGGTCCG CAGGGCCCGC GGGGTGACAA AGGTGAAACG	300
٠	GGCGAACAGG GCGACCGTGG CATTAAAGGC CACCGTGGCT TCAGCGGCCT GCAGGGTCCA	360

	CCGGGCCCGC CGC	GGCAGTCC GGGTG	AACAG GGTCCG	CCG GAGCCAGC	GG GCCGGCGGC	420
5	CCACGCGGTC CGC	CCGGGCAG CGCGG	GCGCG CCGGGC	AAG ACGGTCTG	AA CGGTCTGCCG	480
10.	GGCCCGATCG GCC	CCGCCGGG CCCAC	GCGGC CGCACCC	GTG ATGCGGGT	CC GGTGGGTCCC	540
ক	cceecccc cec	GGCCCGCC AGGCC	CGCCG GGACCG	CGA GCGCGGGT	TT CGACTTCAGC	600
15	TTCCTGCCGC AGO	CCGCCGCA GGAGA	AAGCG CACGACO	GCG GTCGCTAC	FA CCGTGCGTAA	660
20	(2) INFORMATIO	ON FOR SEQ ID	NO:44:			
25	(A) (B)	ENCE CHARACTER LENGTH: 219 a TYPE: amino a STRANDEDNESS:	mino acids cid			•
30 .	(D)	TOPOLOGY: unk	nown			
	(ii) MOLEC	CULE TYPE: pep	tide			
35	(xi) SEQUE	ENCE DESCRIPTI	ON: SEQ ID NO): 44:		
•	Met Gly I	Pro Pro Gly Le	u Ala Gly Pro	Pro Gly Glu	Ser Gly Arg Glu	ı
40]	1	5		10	15	
	Gly Ala F	Pro Gly Ala Gl	u Gly Ser Pro	Gly Arg Asp	Gly Ser Pro Gly	,
4 5		20	25		30	
		Gly Asp Arg Gl	y Glu Thr Gl _}	Pro Ala Gly	Pro Pro Gly Ala	ı
50	Pro Gly A	Ala Pro Gly Al	a Pro Gly Pro	Val Gly Pro	Ala Gly Lys Ser	.
	50		55	60		

		Gly	Asp	Arg	Gly	Glu	Thr	Gly	Pro	Ala	Gly	Pro	Ala	Gly	Pro	Val	Gly
5	•	65					70					75					80
		Pro	Ala	Glv	Ala	Arg	Glv	Pro	Ala	Glv	Pro	Gln	Gly	Pro	Arq	Glv	Asp
10		710		 1		85	,			2	90		•			95	
		Lys	Gly	Glu	Thr	Gly	Glu	Gln	Gly	Asp	Arg	Gly	Ile	Lys	Gly	His	Arg
15					100					105					110		
		Glv	Phe	Ser	Glv	Leu	Gln	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Ser	Pro	Gly
		V-1		115	•			•	120		•			125			•
20																	
		Glu		Gly	Pro	Ser	Gly		Ser	Gly	Pro	Ala		Pro	Arg	Gly	Pro
25			130					135					140				
		Pro	Glv	Ser	Ala	Gly	Ala	Pro	Gly	Lys	Asp	Gly	Leu	Asn	Gly	Leu	Pro
		145	•				150		-	-	_	155					160
 30.																	
		Gly	Pro	Ile	Gly		Pro	Gly	Pro	Arg		Arg	Thr	Gly	Asp		Gly
						165					170					175	
35		Pro	Val	Glv	Pro	Pro	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Pro
				•	180		-			185					190		
;																	
40.		Pro	Ser			Phe	qaA	Phe		Phe	Leu	Pro	Gln		Pro	Gln	Glu
				195					200					205			
45		Lys	Ala	His	Asp	Gly	Gly	Arg	Tyr	Tyr	Arg	Ala					
		-	210	ı				215									
50:	(2)	INFO	RMAT	NOI	FOR	SEQ	ID N	O:45	:								
		(i)	SEC	UENC	E CH	IARAC	TERI	STIC	s:								
		. – •				1: 62											
55																	

(B) TYPE: nucleic acid

	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	
10·	(ii) MOLECULE TYPE: cDNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:	
15	ATGGGCTCTC CGGGTGTTAA CGGCGCCCCT GGTGAAGCGG GCCGCGACGG CAATCCGGGC	60
20 [.]	AACGATGGTC CGCCGGGTCG TGATGGTCAG CCGGGTCACA AAGGTGAGCG TGGCTACCCG	120
	GGTAACATCG GTCCGGTTGG TGCGGCCGGC GCTCCGGGTC CGCACGGTCC GGTAGGCCCA	180
25	GCCGGCAAAC ACGGTAACCG TGGTGAAACG GGTCCGTCCG GTCCGGTAGG TCCGGCGGGT	240
30 ⁻	GCTGTTGGTC CACGCGGCCC GTCCGGCCCG CAGGGTATTC GCGGTGACAA AGGCGAACCG	300
	GGCGAAAAAG GTCCGCGTGG TCTGCCGGGC CTTAAGGGCC ACAACGGTCT GCAAGGTCTG	360
35	CCGGGTATCG CGGGTCACCA CGGTGATCAG GGTGCTCCGG GTTCCGTTGG TCCGGCCGGT	420
	CCGCGTGGCC CGGCTGGTCC GTCTGGTCCG GCCGGTAAAG ACGGCCGTAC GGGCCACCCG	480
40 [.]	GGTACGGTGG GTCCGGCCGG CATTCGCGGT CCGCAAGGTC ACCAGGGTCC GGCGGGTCCG	540
45	CCGGGTCCGC CGGTCCGCC GGTGTTAGCG GTGGCGGTTA TGATTTTGGT	600
	TATGACGGTG ATTTCTATCG TGCGTAA	627
50·		
,		
55		

5	(2)	INFOR	1,1W I I	.ON F	OK 3	EQ I	D NO	. 40.									
:		(i)	SEQU	ENCE	CHA	RACT	ERIS	TICS	:		•						
			(A)	LEN	IGTH :	219	ami	.no a	cids	l							
10			(B)	TYP	E: a	mino	aci	.d									
			(C)	STR	LANDE	DNES	S: 8	ingl	.e								
			(D)	TOP	OLOG	3Y: u	nkno	wn									
15																	
		(ii)	MOLE	CULE	TYP	E: p	epti	de									
4																	
20		(xi)	SEQU	JENCI	E DES	CRIE	PTION	I: SI	EQ II) NO:	46:						
		Met	Gly	Pro	Pro	Gly	Leu	Ala	Gly	Pro	Pro	Gly	Glu	Ser	Gly	Arg	Glu
		1				5					10					15	
25																	
		Gly	Ala	Pro	Gly	Ala	Glu	Gly	Ser	Pro	Gly	Arg	Asp	Gly	Ser	Pro	Gly
					20					25					30		
30																	
5		Ala	Lys	Gly	Asp	Arg	Gly	Glu	Thr	Gly	Pro	Ala	Gly	Pro	Pro	Gly	Ala
				35					40					45			
35				_						_			_			_	_
		Pro	_	Ala	Pro	Gly	Ala		Gly	Pro	Val	Gly		Ala	Gly	Lys	Ser
			50				,	55					60				
40		43	•	•	a 1	61	ent	61	D		01	D	31 -	~ 1	D	17-1	a 1
•		-	qsA	Arg	GIA	GIU		GIÀ	PIO	AIA	GIY		ALA	GIĀ	Pro	Val	
-		65					70					75					80
45		Pro	212	Glv	Δla	Δνα	Glv	Pro	Δla	Glv	Pro	Gln	Glv	Pro	Ara	Gly	Aan
		PIO	AIG	Gry	,,,,,	85	01,	110	A14	GLY	90	3111	O.J		AL 9	95	nop
						-					70						
50		Lva	Glv	Glu	Thr	Glv	Glu	Gln	Glv	Asp	Aro	Glv	Ile	Lvs	Glv	His	Ara
		2,3	1		100	4		- 	1	105		2		_,,	110		· J
					-										_ = -		

		Gly	Phe	Ser	Gly	Leu	Gln	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Ser	Pro	Gly
5				115					120					125			
•																	
		Glu	Gln	Gly	Pro	Ser	Gly	Ala	Ser	Gly	Pro	Ala	Gly	Pro	Arg	Gly	Pro
10			130					135					140				
<u> </u>																	
		Pro	Gly	Ser	Ala	Gly	Ala	Pro	Gly	Lys	Asp	Gly	Leu	Asn	Gly	Leu	Pro
15		145					150					155					160
•		Gly	Pro	Ile	Gly	Pro	Pro	Gly	Pro	Arg	Gly	Arg	Thr	Gly	Asp	Ala	Gly
20						165					170					175	
•		Pro	Val	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Pro		Gly	Pro
. 25					180					185					190		
														_	_		_
•		Pro	Ser		Gly	Phe	Asp	Phe		Phe	Leu	Pro	Gln		Pro	Gln	Glu
				195					200					205			
30				1	_		_,	_	_	_	•	••-					
•		Lys	Ala		Asp	GIÀ	GTÅ		ıyr	Tyr	Arg	Ala					
			210					215									
35	(0)	******	DM T	TOM	EOD	CEO	TD N	0.47									
5	(2)	INFO	KMAI	ION	FOR	SEQ	TD W	0:47	•								•
		(4)	SEQ	ITENC	R CH	מפמ	TERT	STIC	g.								
40		(1)			ngth												
			•	•	PE:			-									
٠			•	-	RAND												
45			-		POLO			_									
			,_			•											
		(ii)	MOL	ECUL	E TY	PE:	CDNA	,									
50																	

³ 5	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:	
	GGAATTCTCC CATGGGCCCG CCGGGTCTGG CGGGCCCTCC GGGTGAAAGC GGTCGTGAAG	60
² 10	GCGCGCCGGG TGCCGAAGGC AGCCCAGGCC GCGAC	95
	(2) INFORMATION FOR SEQ ID NO:48:	
15	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 97 hase pairs	
	(B) TYPE: nucleic acid	
20	(C) STRANDEDNESS: single	
	(D) TOPOLOGY: linear	•
25	(ii) MOLECULE TYPE: cDNA	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:	
30	CTTCCGTCGG GTCCGGCGCT GCCATCGGGC CCCCGGTTTC CCCTAGCACC ACTTTGGCCG	60
35	GGCCGCCCGG GGGGCCCACG TGGCATTATT CGAACCC	97
	(2) INFORMATION FOR SEQ ID NO:49:	
	(i) SEQUENCE CHARACTERISTICS:	
40	(A) LENGTH: 91 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: single	
45	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: cDNA	
50		

147

. 55

5	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:	
J	GGAATTCGGT GCACCGGGCG CGCCGGGTGC CCCAGGCCCG GTGGGCCCGG CGGGCAAAAG	60
10	CGGTGATCGT GGCGAGACCG GTCCGGCGGG C	91
ċ.	(2) INFORMATION FOR SEQ ID NO:50:	
15 20	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 91 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
25	(ii) MOLECULE TYPE: cDNA	
30	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:	
7	CTCTGGCCAG GCCGCCCGGG CCGGCCAGGC CACCCGGGTC GCCCGCGGGC ACCGGGCCGG	60
35	CCAGGCGTCC CGGGCGCCAT TATTCGAACC C	91
40·	Claims 4. A method of producing on Entropollular Matrix Protein (EMP) or fragment thereof conclude of providing a	o oolf on
45	A method of producing an Extracellular Matrix Protein (EMP) or fragment thereof capable of providing a gregate in a cell which does not ordinarily hydroxylate proline comprising Providing a gueloic acid sequence encoding the EMP or fragment thereof which has been entimized.	
50	providing a nucleic acid sequence encoding the EMP or fragment thereof which has been optimized pression in the cell by substitution of codons preferred by the cell for naturally occurring codons not by the cell; incorporating the nucleic acid sequence into the cell; providing hypertonic growth media containing at least one amino acid selected from the group con	preferred
	trans-4-hydroxyproline and 3-hydroxyproline; and contacting the cell with the growth media wherein the at least one amino acid is assimilated into the incorporated into the EMP or fragment thereof.	_
55 _.	2. A method of producing an Extracellular Matrix Protein (EMP) or fragment thereof according to claim 1 wh EMP is selected from the group consisting of human collagen, fibrinogen, fibronectin and collagen-like p	
	3. A method of producing an Extracellular Matrix Protein (EMP) or fragment thereof according to claim 1 or 2	, wherein

the cell is a prokaryote.

5.

10

15

20

25.

30

35.

45

- 4. A method of producing an Extracellular Matrix Protein (EMP) or fragment thereof according to claim 3, wherein the prokaryote is E. coli.
- A method of producing an Extracellular Matrix Protein (EMP) or fragment thereof according to any of claims 2 4, wherein the human collagen is Type I (α1).
- A method of producing an Extracellular Matrix Protein (EMP) or fragment thereof according to claim 5, wherein the nucleic acid encoding human collagen Type I (α1) includes the sequence shown in SEQ.ID.NO.19.
- A method of producing an Extracellular Matrix Protein (EMP) or fragment thereof according to any of claim 2 to 4, wherein the human collagen is Type I (α2).
- 8. A method of producing an Extracellular Matrix Protein (EMP) or fragment thereof according to claim 7, wherein the nucleic acid encoding human collagen Type I (α2)= includes the sequence shown in SEQ.ID.NO.31.
- 9. A method of producing an Extracellular Matrix Protein (EMP) or fragment thereof according to any of claims 1 to 8, wherein the nucleic acid encoding the EMP includes the sequence shown in SEQ.ID.NO. 43.
- A method of producing an Extracellular Matrix Protein (EMP) or fragment thereof according to any of claims 1 to 8, wherein the nucleic acid encoding the EMP includes the sequence shown in SEQ.ID.NO. 46.
- 11. A method of producing an Extracellular Matrix Protein (EMP) or fragment thereof according to any of claims 1 to 10, wherein the nucleic acid sequence includes nucleic acid encoding a physiologically active peptide.
 - 12. A method of producing an Extracellular Matrix Protein (EMP) or fragment thereof according to claim 11, wherein the physiologically active peptide is selected from the group consisting of bone morphogenic protein, transforming growth factor-β and decorin.
 - 13. A method of producing an Extracellular Matrix Protein (EMP) or fragment thereof according to any of claims 1 to 4, wherein the EMP or fragment thereof is a collagen-like peptide.
- 14. A method of producing an Extracellular Matrix Protein (EMP) or fragment thereof according to claim 13, wherein the EMP or fragment thereof includes the amino acid sequence depicted in SEQ.ID.NO. 4.
- 15. A method of producing an Extracellular Matrix Protein (EMP) or fragment thereof according to claim 13, wherein the EMP includes the amino acid sequence depicted in SEQ.ID.NO.40.
- 40 16. A method of producing an Extracellular Matrix Protein (EMP) or fragment thereof according to claim 1, wherein the EMP includes the amino acid sequence depicted in SEQ.ID.NO. 44.
 - 17. A method of producing an Extracellular Matrix Protein (EMP) or fragment thereof according to claim 1, wherein the EMP is a collagen fragment including the amino acid sequence depicted in SEQ.ID.NO. 26.
 - 18. A method of producing an Extracellular Matrix Protein (EMP) or fragment thereof according to claim 1, wherein the EMP is a collagen fragment including the amino acid sequence depicted in SEQ.ID.NO. 46.
 - 19. Nucleic acid encoding a chimeric protein comprising a domain from a physiologically active peptide and a domain from an Extracellular Matrix Protein (EMP) which is capable of providing a self-aggregate.
 - 20. Nucleic acid encoding a chimeric protein according to claim 19, wherein said EMP is selected from the group consisting of human collagen, fibrinogen, fibri
- 21. Nucleic acid encoding a chimeric protein according to claim 19 or 20 wherein said domain from a physiologically active peptide is selected from the group consisting of bone morphogenic protein, transforming growth factor β and decorin.

- 22. Nucleic acid encoding a chimeric protein according to any of claims 19 21, wherein said chimeric protein includes the sequence shown in SEQ.ID.NO.6.
- 23. Nucleic acid encoding a chimeric protein according to any of claims 19 21, wherein said chimeric protein includes the sequence shown in SEQ.ID.NO.8.
- 24. Nucleic acid encoding a chimeric protein according to any of claims 19 21, wherein said chimeric protein includes the sequence shown in SEQ.ID.NO.11.
- 25. Nucleic acid encoding a chimeric protein according to any of claims 19 21, wherein said chimeric protein includes the sequence shown in SEQ.ID.NO.10.
 - 26. A cloning vector comprising nucleic acid according to any of claims 19 21.
- 27. A cloning vector according to claim 26 wherein said cloning vector is selected from the group consisting of plasmid, phage, cosmid and artificial chromosome.
 - 28. A cell transformed by a vector according to claim 26 or 27.

5

25

40

45

- 20 29. A chimeric protein comprising a domain from a physiologically active peptide and a domain from an Extracellular Matrix Protein (EMP) which is capable of providing a self-aggregate.
 - **30.** A chimeric protein according to claim 29 wherein said EMP is selected from the group consisting of human collagen, fibrinogen, fibronectin and collagen-like peptide.
 - 31. A chimeric protein according to claim 29 or 30 wherein said domain from a physiologically active peptide is selected from the group consisting of bone morphogenic protein, transforming growth factor β and decorin.
- **32.** A chimeric protein according to any of claims 29 31, wherein said chimeric protein includes the sequence shown in SEQ.ID.NO.6.
 - 33. A chimeric protein according to any of claims 29 31, wherein said chimeric protein includes the sequence shown in SEQ.ID.NO.8.
- 35 **34.** A chimeric protein according to any of claims 29 31, wherein said chimeric protein includes the sequence shown in SEQ.ID.NO.10.
 - 35. A chimeric protein according to any of claims 29 31, wherein said chimeric protein includes the sequence shown in SEQ.ID.NO.11.
 - 36. Human collagen or fragment thereof produced by a prokaryotic cell, the human collagen or fragment thereof being capable of providing a self-aggregate.
 - 37. Human collagen or fragment thereof produced by a prokaryotic cell according to claim 36 wherein the human collagen or fragment thereof is encoded for by nucleic acid having the sequence shown in SEQ.ID.NO.19.
 - 38. Human collagen or fragment thereof produced by a prokaryotic cell according to claim 36 wherein the human collagen or fragment thereof is encoded for by nucleic acid having the sequence shown in SEQ.ID.NO.39.
- 39. Human collagen or fragment thereof produced by a prokaryotic cell according to claim 36 wherein the human collagen or fragment thereof is encoded for by nucleic acid having the sequence shown in SEQ.ID.NO.43.
 - **40.** Human collagen or fragment thereof produced by a prokaryotic cell according to claim 36 wherein the human collagen or fragment thereof is encoded for by nucleic acid having the sequence shown in SEQ.ID.NO.45.
 - 41. Human collagen or fragment thereof produced by a prokaryotic cell according to claim 36 wherein the collagen or fragment thereof is encoded for by nucleic acid having the sequence shown in SEQ.ID.NO.31.

- 42. Nucleic acid comprising the sequence shown in SEQ.ID.NO. 19.
- 43. Nucleic acid comprising the sequence shown in SEQ.ID.NO. 31.
- 5 44. Nucleic acid comprising the sequence shown in SEQ.ID.NO. 43.

10

15

20

25

30

35

40

45

50

55

- 45. Nucleic acid comprising the sequence shown in SEQ.ID.NO. 45.
- 46. Nucleic acid encoding a human Extracellular Matrix Protein (EMP) or fragment thereof wherein the codon usage in the nucleic acid sequence reflects preferred codon usage in a prokaryotic cell.
- 47. Nucleic acid according to claim 46 wherein the prokaryotic cell is *E. coli*.
- 48. Nucleic acid according to claim 43 wherein the EMP is selected from the group consisting of collagen, fibrinogen, fibronectin and collagen-like peptide.

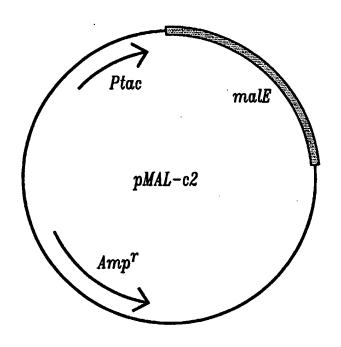
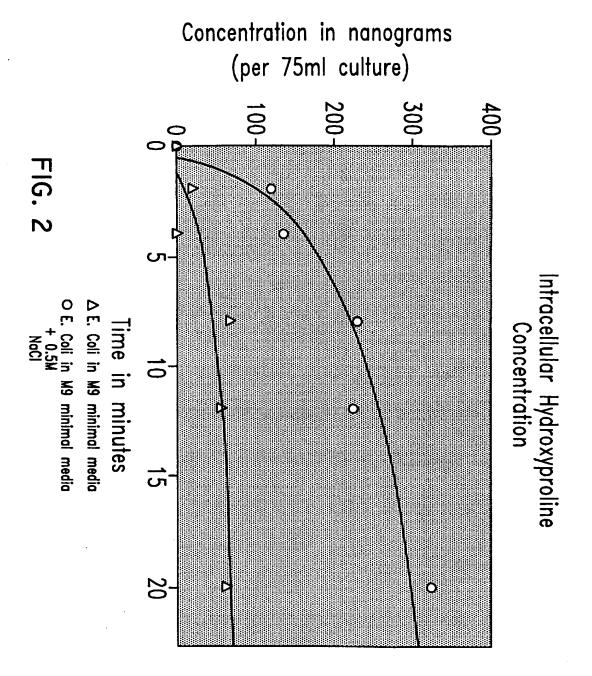


FIG. I



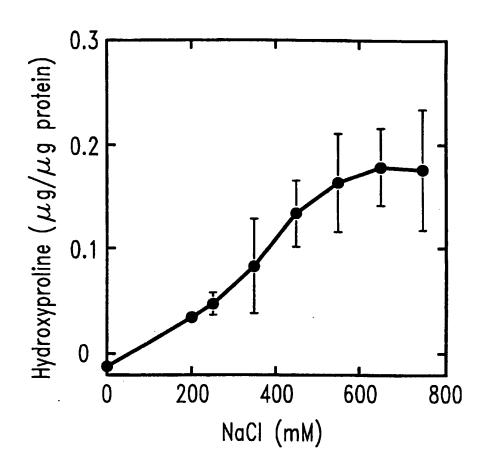


FIG. 2A

5'- CAGCTGTCTT ATGGCTATGA TGAGAAATCA ACCGGAGGAA TTTCCGTGCC TEGECECATE SETECETETE GIVETEGTES TETEVETESE CECECTEGTS CACCTGGTCC CCAAGGCTTC CAAGGTCCCC CTGGTGAGCC TGGCGAGCCT GGAGCTTCAG GTCCCATGGG TCCCCGAGGT CCCCCAGGTC CCCCTGGAAA GAATGGAGAT GATGGGGAAG CTGGAAAACC TGGTCGTCCT GGTGAGCGTG GCCTCCTGG GCCTCAGGGT GCTCGAGGAT TGCCCGGAAC AGCTGGCCTC CCTGGAATGA AGGGACACAG AGGTTTCAGT GGTTTGGATG GTGCCAAGGG AGATGCTGGT CCTGCTGGTC CTAAGGGTGA GCCTGGCAGC CCTGGTGAAA ATGGAGCTCC TGGTCAGATG GGCCCCGTG GCCTGCCTGG TGAGAGAGGT CGCCCTGGAG CCCCTGGCCC TGCTGGTGCT CGTGGAAATG ATGGTGCTAC TEGTECTECC GESCUCCTG GTOCCACCGG CCCGGCTGGT CCTCCTGGCT TCCCTGGTGC TGTTGGTGCT AAGGGTGAAG CTGGTCCCCA AGGGCCCCCA GECTCTGAAG GTCCCCAGGG TGTGCGTGGT GAGCCTGGCC CCCCTGGCCC TECTEGTECT CCTEGCCCTG CTGCAAACCC TGGTGCTGAT GCACAGCCTG GTGCTAAAGG TGCCAATGGT GCTCCTGGTA TTGCTGGTGC TCCTGGCTTC CCTGGTGCCC GAGGCCCCTC TGGACCCCAG GGCCCCGGGG GCCCTCCTGG TCCCAAGGGT AACAGCGGTG AACCTGGTGC TCCTGGCAGC AAACGAGACA CTCGTCCTAA GGCAGACCCT GGCCCTGTTG GTGTTCAAGG ACCCCCTGGC CCTECTEGAG AGGAAGGAAA GCGAGGAGCT CGAGGTGAAC CCGGACCCAC TEGECETECCE GEACCECETG GEGACEGTEG TEGACETEGT ACCEGTEGTT TCCCTGGCGC AGATGGTGTT GCTGGTCCCA AGGGTCCCGC TCGTGAACGT CGTTCTCCTG GCCCCGCTGG CCCCAAAGGA TCTCCTGGTG AAGCTGGTCG TCCCGGTGAA GCTGGTCTGC CTGGTGCCAA GGGTCTGACT GGAAGCCCTG CCACCCTGG TCCTGATGGC AAAACTCGCC CCCCTGGTCC CGCCGGTCAA CATGGTCGCC COGGACCCCC AGGCCCACCT GGTGCCCGTG GTCAGGCTGG TGTCATCCCA TTCCCTCCAC CTAAACGTCC TCCTGCACAG CCCGGCAAGG CTGGAGAGGG AGGTGTTCCC GGACCCCCTG GCGCTGTCGG TCCTGCTGGC AAAGATGGAG AGGCTGGAGC TCAGGGACCC CCTGGCCCTG CTGGTCCCGC TGGCGAGAGA GGTGAACAAG GCCCTGCTGG CTCCCCGGA TTCCAGGGTC TCCCTGGTCC TGCTGGTCCT CCAGGTGAAG CAGGCAAACC TGGTGAACAG GGTGTTCCTG CAGACCTTGG CGCCCCTGCC CCCTCTGGAG CAAGAGGCCGA CACACGTTTC CCTCCCCACC GTCGTGTCCA ACGTCCCCCT GGTCCTGCTG CACCCCGACG GGCCAACGGT GCTCCCGGCA ACGATGGTGC TAAGGGTGAT CCTGGTGCCC CTGGAGCTCC CGGTAGCCAG GGCGCCCCTG GCCTTCAGGG AATGCCTGGT GAACGTGGTG CACCTGGTCT TCCAGGGCCCT AAGGGTGACA CAGGICATIC TEGTECCAAA GGTGCTGATG GCTCTCCTGG CAAAGATGGC

FIG. 3A

GICCGICGIC TGACCGECCC CAITGGICCI CCTCCCCCIG CIGGICCCCC
TGGTGACAAG GGTGAAAGTG GTOCCAGGGG CCCTGCTGGT CCCACTGGAG
CTCGTCGTCC CCCCCGAGAC CGTCGTGAGC CTCGTCCCCC CCCCCCTCCT
GCCTTTCCTG GCCCCCCTGG TGCTGACGGC CAACCTGGTG CTAAAGGCGA
ACCTGGTGAT GCTGGTGCCA AAGGCGATGC TGGTCCCCCT GGGCCTGCCG
GACCCCCTCG ACCCCCTCGC CCCATTCGTA ATGTTCGTCC TCCTCCACCC
AAAGGTGCTC GOGGCAGCGC TGGTCCCCCT GGTGCTACTG GTTTCCCTGG
TECTECTEGE CGAGTEGGTC CTCCTGGCCC CTCTGGAAAT GCTGGACCCC
CTGGCCCTCC TGGTCCTGCT GGCAAAGAAG GCGGCAAAGG TCCCCGTGGT
GAGACTGGCC CTGCTGGACG TCCTGGTGAA GTTGGTCCCC CTGGTCCCCC
TECCCCTECT CECCAGAAAG CATCCCCTEG TECTEATEGT CCTECTEGTG
CTCCTGGTAC TCCCGGGCCT CAAGGTATTG CTGGACAGCG TGGTGTGGTC
GCCCTGCCTG GTCAGAGAGG AGAGAGAGGC TTCCCTGGCCC
CTCTCGTGAA CCTGGCAAAC AAGGTCCCTC TGGAGCAAGT GGTGAACGTG
GTCCCCCCGG TCCCATGGGC CCCCTGGAT TGGCTGGACC CCCTGGTGAA
TCTGGACGTG AGGGGGCTCC TGCTGCCGAA GGTTCCCCTG GACGAGACGG
TTCTCCTGGC GCCAAGGGTG ACCGTGGTGA GACCGGCCCC GCTGGACCCC
CTEGTECTCC TEGTECTCCT GETECCCCTG GCCCCGTTGG CCCTGCTGGC
AACAGTEGTG ATCGTEGTEA CACTEGTECT CCTGGTCCCG CCCGTCCCGT
CCCCCCCCCT CCCCCCCCTG CCCCCCCCG ACCCCAACCC CCCCCTGGTG
ACAAGGGTGA GACAGGGGAA CAGGGGGACA GAGGCATAAA GGGTCACCGT
GGCTTCTCTG GCCTCCAGGG TCCCCCTGGC CCTCCTGGCT CTCCTGGTGA
ACAACGTCCC TCTGGAGCCT CTGGTCCTGC TCGTCCCCCA GGTCCCCCTG
SCICICCICG TECTOCICCO AAACATGCAC TOAACCGICT COCTGGCCCC
ATTEGECCCC CIEGROCICE OFFICECACT CETEATECIE GICCIETTES
ICCCCCCGGC CCICCTGGAC CTCCTGGTCC CCCTGGTCCT CCCAGCGCTG
STITICGACIT CAGCITCCTC CCCCAGCCAC CTCAAGAGAA GGCTCACGAT
GTGGCCGCT ACTACCGGGC T-3'

FIG. 3B

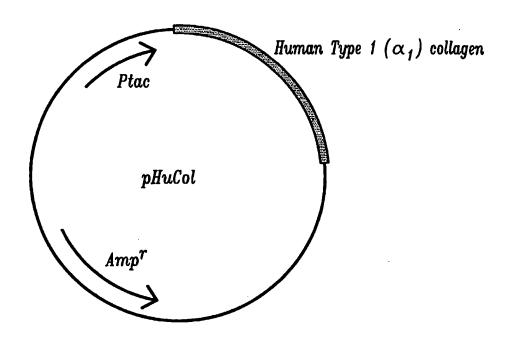


FIG. 4

5'- CAGCTGTCTT ATGCTATGA TGAGAAATCA ACCGGAGGAA TTTCCGTGCC
TGGCCCCATG GGTCCCTCTG GTCCTCGTG TCTCCCTGGC CCCCCTGGTG
CACCTGGTCC CCAAGGCTTC CAAGGTCCCC CTGGTGAGCC TGGCGAGCCT
GGAGCTTCAG GTCCCATGGG TCCCCGAGGT CCCCCTGGAAA
GAATGGAGAT GATGGGGAAG CTGGAAAACC TGGTCGTCCT-3'

FIG. 5

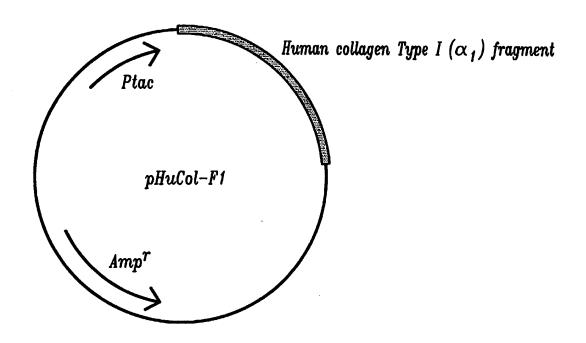


FIG. 6

GGA TCC ATG GGG CTC GCT GGC CCA CCG GGC GAA CCG GGT CCG CCA GGC CCG AAA GGT CCG CGT GGC GAT AGC GGG CTC CCG GGC GAT TCC TAA TGG ATC C

FIG. 7

Gly-Leu-Ala-Gly-Pro-Pro-Gly-Glu-Pro-Gly-Pro-Pro-Gly-Pro-Lys-Gly-Pro-Arg-Gly-Asp-Ser

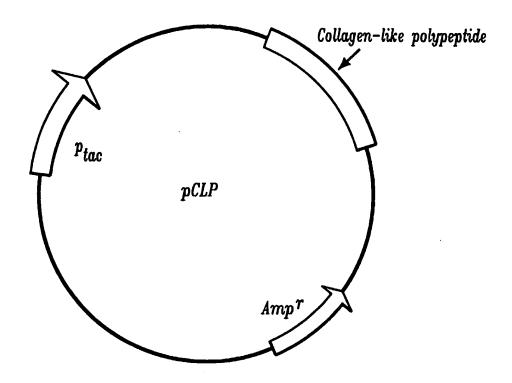


FIG. 9

5'-	CAGCGGGCCA	GGAAGAAGAA	TAAGAACTGC	CGGCGCCACT	CGCTCTATGT
	GGACTTCAGC	GATGTGGGCT	GGAATGACTG	GATTGTGGCC	CCACCAGGCT
	ACCAGGCCTT	CTACTGCCAT	GGGGACTGCC	CCTTTCCACT	GGCTGACCAC
	CTCAACTCAA	CCAACCATGC	CATTGTGCAG	ACCCTGGTCA	ATTCTGTCAA
	TTCCAGTATC	CCCAAAGCCT	GTTGTGTGCC	CACTGAACTG	AGTGCCATCT
	CCATGCTGTA	CCTGGATGAG	TATGATAAGG	TGGTACTGAA	AAATTATCAG
		TAGAGGGATG			

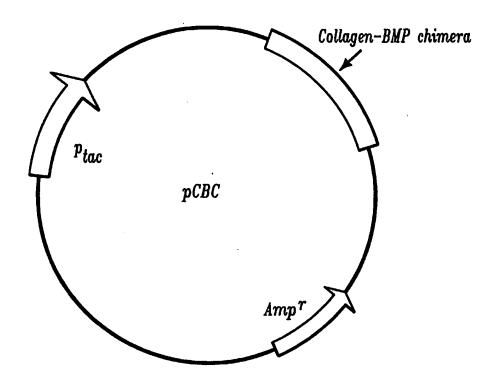
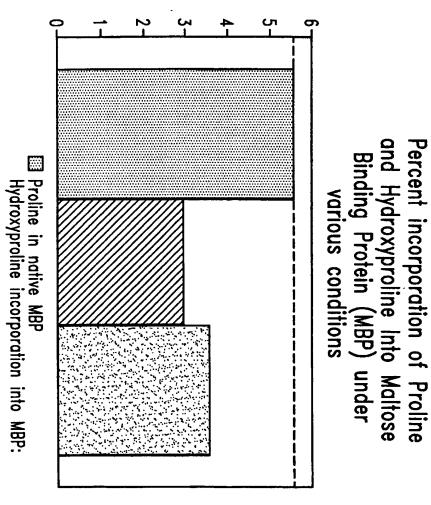


FIG. II

Mole percent of MBP



☐ Grown in isotonic media

🖾 Grown in hypertonic media

IG. 12

10	20	30	40	SO	60
QLSYGYDEKS	TGGISVPGPM	GPSGPRGLPG	PPGAPGPQGF	OGPPGEPGEP	Gasgpmgprg
70	80	90	100	110	120
PPGPPGKNGD	DGEAGXPGRP	GergppgpQG	ARGLPGTAGL	PGMKGHRGFS	GLDGAKGDAG
130	140	150	160	170	180
PAGPKGEPGS	PGENGAPGQM	GPRGLPGERG	RPGAPGPAGA	RONDGATGAA	GPPGPTGPAG
190	200	210	220	230	240
PPGFPGAVGA	KGEAGPQGPR	GSEGPQGVRG	EPGPPGPAGA	AGPAGNPGAD	GQPGAXGANG
250	260	270	280	290	300
APGIAGAPGF	PGARGPSGPQ	GPGGPPCPKG	NSGEPGAPGS	KGDTGAKGEP	GPVGVQGPPG
310	320	330	340	350	360
PAGEEGKRGA	RGEPGPTGLP	GPPGERGGPG	SRGFPGADGV	Agpkgpager	GSPGPAGPKG
370	380	390	400	410	420
SPGEAGRPGE	AGLPGAKGLT	GSPGSPGPDG	KTGPPGPAGQ	DGRPGPPGPP	Gargoagymg
430 FPGPKGAAGE					480 GEQGPAGSPG
490	500	510	520	530	540
FOGLPGPAGP	PGEAGKPGEQ	GVPGDLGAPG	PSGARGERGF	PGERGVQGPP	GPAGPRGANG
					GADGSPGKDG
610		630	640	650	660
Vagltgpigp		GESGPSGPAG	PTGARGAPGD	RGEPGPPGPA	GFAGPPGADG
670		690	700	710	720
QPGAKGEPGD		GPAGPAGPPG	PIGNVGAPGA	KGARGSAGPP	GATGFPGAAG
730	740	750	760	770	780
RVGPPGPSGN	Agppgppgpa	GKEGGKGPRG	ETGPAGRPGE	VGPPGPPGPA	GEKGSPGADG
790	800	810	820	830	840
PAGAPGTPGP	QGIAGQRGVV	GLPGORGERG	FPGLPGPSGE	PGKQGPSGAS	CERGPPGPHG
850	860	870	880	890	900
PPGLAGPPGE	Sgregapaae	GSPGRDGSPG	AKGDRGETGP	AGPPGAXGAX	Gapgpvgpag
910	920	930	940	950	960
KSGDRGETGP	AGPAGPVGPA	Gargpagpog	PRGDKGETGE	QGDRGIKGHR	GFSGLQG?PG
970	.980	990		1010	1020
PPGSPGEQGP	SGASGPAGPR	GPPGSAGAPG		IGPPGPRGRT	GDAGPVGPPG
1030	1040	1050	1060	1070	1080
PPGPPGPPGP	PSAGFDFSFL	PQPPQEXAHD	GGRYYRARSQ	RARKKNIKNCR	RHSLYVDFSD
0901 Pavikdradv			1120 TOVIAHNTZN	1130 LVNSVNSSIP	1140 KACCVFTELS
1150 AISMLYLDEY	1160 DKVVLKNYQE	1170 MVVEGCGCR*	.1180	1190	1200

FIG. 13

10	20	30	40	. 50	60 CCGGAGGAAT
					120
TTCCGTGCCT	GGCCCCATGG	GTCCCTCTGG	TCCTCGTGGT	CTCCCTGGCC	CCCCTGGTGC
					180 GAGCTTCAGG
190 TCCCATGGGT	200 CCCCGAGGTC	210 CCCCAGGTCC	220 CCCTGILAAAG	230 AATGGAGATG	240 ATGGGGAAGC
	260 GGTCGTCCTG				300 CTCGAGGATT
310 GCCCGGAACA	320 GCTGGCCTCC	330 CTGGAATGAA	340 GGGAC.\CAGA	350 GCTTTCAGTG	360 GTTTGGATGG
370 TGCCAAGGGA	380 GATGCTGGTC	390 CTGCTGGTCC	400 TAAGGTTGAG	410 CCTGGCAGCC	420 CTGGTGAAAA
430 TGGAGCTCCT	.440 GGTCAGATGG	450 GCCCCCGTGG	460 CCTGCCTGGT	470 GAGAGAGGTC	480 GCCCTGGAGC
					540 GGCCCCCTGG
	560 CCCGCTGGTC				600 AGGGTGAAGC
	620 GGGCCCCGAG	630 GCTCTGAAGG	640 TCCCC.4GGGT	650 GTGCGTGGTG	. 660 AGCCTGGCCC
670 CCCTGGCCCT	680 GCTGGTGCTG	690 CTGGCCCTGC	700 TGGAŁ\CCCT	710 GGTGCTGATG	720 GACAGCCTGG
730 TGCTAAAGGT	740 GCCAATGGTG	750 CTCCTGGTAT	760 TGCTOJTGCT	770 CCTGGCTTCC	780 CTGGTGCCCG
790 AGGCCCCTCT	008 GGACCCCAGG		820 CCCTCCTGGT		840 ACAĠCGGTGA
850 ACCTGGTGCT	860 CCTGGCAGCA	870 AAGGAGACAC	088 Daatedtddt	890 GGAGAGCCTG	900 GCCCTGTTGG
910 TGTTCAAGGA	920 CCCCCTGGCC	930 CTGCTGGAGA	940 GGAAG:3AAAG	950 CGAGGAGCTC	960 GAGGTGAACC
970 CGGACCCACT	980 GGCCTGCCCG	990 GACCCCCTGG		1010 GGACCTGGTA	1020 GCCGTGSTTT
1030 CCCTGGCGCA	1040 GATGGTGTTG	1050 CTGGTCCCAA	1060 GGGTC:CGCT	1070 CSTGAACGTG	1080 GTTCTCCTGG
1090 CCCCGCTGGC	1100 CCCAAAGGAT	1110 CTCCTGGTGA	1120 AGCTGGTCGT	1130 CCCGGTGAAG	1140 CTGGTCTGCC
1150 TGGTGCCAAG	1160 GGTCTGACTG		1180 CAGCCCTGGT		
1210 CCCTGGTCCC	1220 GCCGGTCAAG	1230 ATGGTCGCCC	CGGACIICCCA	GGCCCACCTG	1260 GTGCCCCTGG

1270	1280 GTGATGGGAT	1290	1300	· 1310	1320
1330	1340 GGTGTTCCCG	CACCCCCTGG	CGCTGTCGGT	CCTGCTGGCA	AAGATGGAGA
1390	1400 CAGGGACCCC	1410	TGGTC::CGCT	GGCGAGAGAG	GTGAACAAGG
			•		
1450	1460 TCCCCCGGAT	1470 TCCAGGGTCT	CCCTG:TCCT	CCTCGTCCTC	CAGGIGAAGC
					1560
1510 AGGCAAACCT	GGTGAACAGG	GTGTTCCTGG	AGACC'ITGGC	COCCOCCC	CCTCTGGAGC
	_				1620
1570 AAGAGGCGAG	AGAGGTTTCC	CTGGCGAGCG	TGGTG'IGCAA	CCTCCCCTC	GTCCTCCTCG
					1680
ACCCCGAGGG	CCCAACGGTG	CTCCCGGCAA	CGATGGTGCT	AAGGGTGATG	CTGGTGCCCC
					1740
TOGAGCTCCC	GGTAGCCAGG	GCGCCCCTGG	CCTTCAGGGA	ATGCCTGGTG	AACGTGGTGC
1750	1760	1770	1780	1790	1800
AGCTGGTCTT	CCAGGGCCTA	AGGGTGACAG	AGGTGATGCT	GGTCCCAAAG	GTGCTGATGG
1810	1820	1830	1840	1850	1860
CTCTCCTGGC	ALAGATGGCG	TCCGTGGTCT	GACCGGCCCC	ATTGGTCCTC	CTGGCCCTGC
1870	1880	1890	1900	1910	1920
	GGTGACAAGG				•
	1940				
regregated					CCTTTCCTCC
1990	2000 GCTGACGGCC				2040
2050 AGGCGATGCT	2060 GGTCCCCCTG	2070 GGCCTGCCGG	2080 ACCCGCTGGA	2090 CCCCCTGGCC	2100 CCATTGGTAA
		•		٠.	
TGTTGGTGCT	2120 CCTGGAGCCA	AAGGTGCTCG	CGGCAGCGCT	CGTCCCCCTG	GTGCTACTGG
	2180	*			
TTTCCCTGGT	GCTGCTGGCC	GAGTCGGTCC	TCCTGSCCCC	TCTGGAAATG	CTGGACCCCC
				:	2280
TGGCCCTCCT	GGTCCTGCTG	GCAAAGAAGG	CGGCAAAGGT	CCCCGTGGTG	AGACTGGCCC
2290	2300	2310	2320	2330	2340
					GCGAGAXAGG
					2400
ATCCCCTGGT	GCTGATGGTC	CTGCTGGTGC	TCCTGGTACT	CCCGGGCCTC	AAGGTATTGC
				2450	
					TCCCTGGTCT
	2480				2520 GTGAACGTGG
1,000,000	1 C 1 C C I CANC	CIOOCANCA	Virter	011000001010	SIGNACOICE

FIG. 14B

2530	2540	2550	2560	• 2570	2580
TCCCCCCGGT	CCCATGGGCC	CCCCTCGATT	CCCTCCACCC	CCTGGTGAAT	CTGGACGTGA
					2640
GGGGGCTCCT	GCTGCCGAAG	GTTCCCCTGG	ACGAGACGGT	TCTCCTGGCG	CCAAGGGTGA
					2700
	•				GTGCCCCTGG
2710	2720	2730	2740	2750	2760
CCCCGTTGGC	CCTGCTGGCA	AGAGTGGTGA	TCGTGGTGAG	ACTOGICCIG	CTGGTCCCGC
					2820
CCGTCCCGTC	GCCCCCCTG	GCGCCCGTGG	CCCCCCCGGA	CCCCAAGGCC	CCCCTGGTGA
2830	2840	2850	2860	2870	2880
CAAGGGTGAG	ACAGGCGAAC	AGGGCGACAG	AGGCATAAAG	GGTCACCGTG	GCTTCTCTGG
2890	2900	2910	2920	2930	2940
					CTGGAGCCTC
2950	2960	2970	2980	2990	3000
TGGTCCTGCT	GGTCCCCGAG	excécerce	CTCTGCTGGT	GCTCCTGGCA	AAGATGGACT
3010	3020	3030	3040	3050	3060
CAACGGTCTC	CCTGGCCCCA	TTGGGCCCCC	TOGTCCTCGC	CCTCCCACTG	GTGATGCTGG
3070	3080	3090	3100	3110	3120
					CCAGCGCTGG
3130	3140	3150	. 3160	3170	3180 GTGGCCGCTA
3190	3200	3210	3220	3230	3240
CIACCGGGCT	agattcccxcc	GGGCCAGGAA	GAAGAATAAG	AACTGCCGGC	GCCACTCGCT
3250	3260	3270	3280	3290	3300
CTATGTGGAC	TTCAGCGATG	TGGGCTGGAA	TGACTGGATT	GTGGCCCCAC	CAGGCTACCA
3310	3320	3330	3340	3350	3360 .
GGCCTTCTAC	TGCCATGGGG	ACTGCCCCTT	TCCACTGGCT	GACCACCTCA	ACTCAACCAA
3370	3380	3390	3400	3410	3420
CCATGCCATT	GIGCAGACCC	TGGTCAATTC	TGTCAATTCC	AGTATCCCCA	Axccctcttc
3430	3440	3450	3460	3470	3480
TGTGCCCACT	GAACTGAGTG	CCATCTCCAT	GCTGTACCTG	CATGAGTATG	ataaggtggt
3490	3500	3510	3520	3530	3540
ACTGAAAAAT	TATCAGGAGA	TCGTAGTAGA	GGGATGTGGG	TGCCGCTAAa	agctt

FIG. 14C

60 Gasgpygprg	50 QGPPGEPGEP	40 PPGAPGPQGF	30 GPSGPRGLPG	20 TGGISVPGPM	. 10 QLSYGYDEKS
120	110	100	00		
180 GPPGPTGPAG	170 RGNDGATGAA	160 RPGAPGPAGA	150 GPRGLPGERG	140 PGENGAPGQM	130 PAGPKGEPGS
240 GQPGAKGANG	230 AGPAGNPGAD	220 EPGPPGPAGA	210 GSEGPQGVRG	200 KGEAGPQGPR	
300 GPVGVQGPPG	290 Kgdtgakgep	280 NSGEPGAPGS	270 GPGGPPGPKG	260 PGARGPSGPO	250 APGIAGAPGF
360 GSPGPAGPXG	350	340	330	, ,,,,	
420 GARGQAGVIIG	410	.400	390	380	220
480 GEQGPAGSPG	470	460	450	440	. 430
540 GPAGPRGANG	530	520	510	500	490
600 GADGSPGKDG	590	580	570	\$60	550
660 GFAGPPGADG	650	640	. 630	620	610
720 GATGFPGAAG	710	700	690	680	670
780 GEKGSPGADG	770	760	750	740	730
840 GERGPPGFMG	830	820	810	. 800	790
900 GAPGPVGPAG	890	880	870	860	850
960 GFSGLQGPPG	· 950	940	930	920	910
1020 GDAGPVGPPG	. 1010	1000	990	980	970
1020 EXPICCVRQLY	1070	1060	1050	1040	1030
	1130	1120	1110	1100	1090
	1190	1180	1170	1160	1150

FIG. 15

10	20	30	40	. 50	60
10 gggaaggatt					
70 TICCGTGCCT	80 GGCCCCATGG	90 GTCCCTCTGG	100 TCCTCGTGGT	110 CTCCCTGGCC	CCCCTGGTGC
	140	150	160	170	180
190 TCCCATGGGT	200 CCCCGAGGTC	210 CCCCAGGTCC	220 CCCTG:LAAAG	230 AATGGAGATG	240 ATGGGGAAGC
250 TGGAAAACCT	260 GGTCGTCCTG	270 GTGAGCGTGG	280 GCCTCCTGGG	290 CCTCAGGGTG	300 CTCGAGGATT
310	320	330	340	350	360 GTTTGGATGG
370	380	390	400	410	420 CTGGTGAAAA
430	440	450	460	470	480 GCCCTGGAGC
490	500	510	520	530	S40 GGCGCCCTGG
550 TCCCACCGGC	560 CCCCCTGGTC	570 CTCCTGGCTT	580 CCCTC/ITGCT	590 GTTGGTGCTA	006 AGGGTGAAGC
610	620	. 630	640	650	660 AGCCTGGCCC
670	680	690	700	710	720 GACAGCCTGG
730 TGCTAAAGGT	740 GCCAATGGTG	750 CTCCTGGTAT	760 TGCTGGTGCT	770 CCTGGCTTCC	780 CTGGTGCCCG
790	800	810	820	830	840 ACAGCGGTGA
850	. 860	870	880	890	900 GCCCTGTTGG
910	920	930	940	950	960 GAGGTGAACC
970	.980	990	1000	. 1010	1020 CCCGTCGTTT
. 1030	1040	1050	1060	. 1070	
1090	1100	1110	1120	1130	1140 CTGGTCTGCC
1150	1160	1170	1180	1190	
1210	1220	1230	1240	1250	

FIG. 16A

1270 1280 1290 1300 1310 1310 1320 1320 1330 1310 131						
1390	1270 TCAGGCTGGT	1280 GTGATGGGAT	1290 TCCCTGGACC	1300 TAAAGGTGCT	1310 GCTGGAGAGC	1320 CCGGCAAGGC
1390	1330	1340	1350	1360	1370	1380
1450						
1510	GGCTGGAGCT	CAGGGACCCC	CTGGCCCTGC	TGGTCCCGCT	GGCGAGAGAG	GTGXACAAGG
AGGCAAACCT GGTGAACAGG GTGTTCCTGG AGACCTTGGC CCCCTGGCG CCCCTGGCGC CCCCTGGCAGC The control of the control	1450 CCCTGCTGGC	1460 TCCCCCGGAT	1470 TCCAGGGTCT	1480 CCCTGGTCCT	1490 GCTGGTCCTC	1500 CAGGTGAAGC
### AGAGGCGAG AGAGGTTTCC CTGGCGAGCG TGGTGTGCAA GGTCCCCTG GTCCTGCTGG 1630	1510 AGGCAAACCT	1520 GGTGAACAGG	1530 GTGTTCCTGG	1540 AGACCTTGGC	1550 GCCCCTGGCC	1560 CCTCTGGAGC
1630 1640 1650 1660 1670 1680 ACCCCGAGGG GCCAACGGTG CTCCCGGCAA CGATGGTGCT AAGGGTGATG CTGGTGCCCC 1690 1700 1710 1720 1730 1740 TGGAGCTCCC GGTAGCCAGG GCGCCCCTGG CCTTCAGGGA ATGCCTGGTG AACGTGGTGC 1750 1760 1770 1780 1790 1890 AGCTGGTCTT CCAGGGCCTA AGGGTGACG AGGTGATGCT GGTCCCAAAG GTGCTGATGG 1810 1820 1830 1840 1850 1860 CTCTCCTGGC AAAGATGGCG TCCGTGGTCT GACCGACCC ATTGGTCCTC CTGGCCCTGC 1870 1880 1890 1900 1910 1920 TGGTGCCCCT GGTGACAAGG GTGAAAGTGG TCCCAGGGC CCTGCTGGTC CCACTGGAGC 1930 1940 1950 1960 1970 1970 1980 TCGTGGTGCC CCCGGGAGACC GTGGTGAGCC TGGTCCCCCC GGCCCTGCTG GCTTTCCTGG 1990 2000 2010 2020 2030 2040 CCCCCCCTGGT GCTGACGGC AACCTGGTGC TAAAGGCGAA CCTGGTGATG CTGGTCCCAA 2050 2060 2070 2080 2090 2100 AGGCGATGCT GGTCCCCCT GGCCTGCCG ACCCGTTGAT CTGGTCCCAA 2050 2060 2070 2080 2090 2100 AGGCGATGCT GGTCCCCCTG GGCCTGCCG ACCCGTTGAT CTGGTCCCAA 2110 2120 2130 2140 2150 CCCCTTGGC CCATTGGTAA 2110 2120 2130 2140 2150 CCCCTTGGC CTGGTCCCCC 2230 2240 2250 2260 2270 2280 CCCCTTGGC CTGGACCCCC 2230 2240 2250 2260 2270 2280 CTGGAAAGAAG CTGGACCCCC 2290 2300 2240 2250 2260 2270 2280 TGCTGGACCTCT GGTCCTCGTG GCAAAGAAG CGGCAAAGAT CCCCCTTGGC GGCCTGCCCC 2270 2300 2240 2250 2260 2270 2280 TGCTGGACGT CCTGGTGAAG TTGGTCCCCC TGGCCCTGG GCCCTGCCCCCCCCCC	1570 AAGAGGCGAG	1580 AGÁGGTTTCC	1590 CTGGCGAGCG	1600 TGGTGTGCAA	1610 GGTCCCCCTG	1620 GTCCTGCTGG
1690 1700 1710 1720 1730 1740 TGGAGCTCCC GGTAGCCAGG GCGCCCCTGG CCTTCAGGGA ATGCCTGGTG AACGTGGTGC 1750 1760 1770 1780 1790 1890 AGCTGGTCTT CCAGGGCCTA AGGGTGACG AGGTGATGCT GGTCCCAAAG GTGCTGATGG AGCTGTCTT CCAGGCCTA AGGGTGACG AGGTGATGCT GGTCCCAAAG GTGCTGATGG 1810 1820 1830 1840 1850 1260 CTCTCCTGGC AAAGATGGG TCCGTGGTCT GACCGXCCC ATTGGTCCT CTGGCCCTGC 1870 1880 1890 1900 1910 1920 TGGTGCCCCT GGTGACAAGG GTGAAAGTGG TCCCAGGGC CCTGCTGGTC CCACTGGAGC 1930 1940 1950 1960 1970 1980 TCGTGGTGCC CCCCGGAGACC GTGGTGAGCC TGGTCCCCC GGCCCTGCT GCTTTCCTGG 1990 2000 2010 2020 2030 2040 CCCCCCTGGT GCTGACGCC AACCTGGTGC TAAAGGCGAA CCTGGTGATG CTGGTCCCAG 2050 2060 2070 2080 2090 2100 AGGCGATGCT GGTCCCCCT GGCCCTCGG ACCCGTTGGA CCCCCTGGCC CCATTGGTACA 2050 2060 2070 2080 2090 2100 AGGCGATGCT GGTCCCCCTG GGCCTGCCG ACCCGTTGGA CCCCCTGGCC CCATTGGTACA 2110 2120 2130 2140 2150 2160 TGTTGGTGCT CCTGGAGCCA AAGGTGCTCG CGGCAXGCT GGTCCCCCTG GTGCTACTGG 2270 2180 2240 2250 2260 2270 2220 TGCCCCTCCTG GGTCCTCGCC GAGTCGGTCC TCCTGGACATG CTGGACCCCC 2230 2240 2250 2260 2270 2270 2280 TGGCCCTCCT GGTCCTGCC GAGTGGTCC TCCTGGACATG CTGGACCCCC 2230 2240 2250 2260 2270 2270 2280 TGCCCCTCGT GCTCCTGCTG GCAAAGAAG CGGCAAAGGT CCCCGTGGTG AGGCTGCCCCC 2270 2300 2240 2250 2260 2270 2280 TGCCCCTCCT GGTCCTCCTG GCAAAGAAGG CGGCAAAGGT CCCCGTGGTG AGGCTGCCCCC 2290 2300 2310 2310 2320 2330 2340 TGCTGGACACC CCTGGTGAAG TTGGTCCCCC TGGTCCCCT GGCCCCTCG GCGAGAAAGACCCC 2250 2350 2360 2370 2380 2390 2400 ATCCCCTGGT GCTGATGGTC CTGCTGGTGC TCCCTGGTACT CCCGGGCCTC AAGGTATTGC 2410 2420 7420 7430 7440 7450 7460 ATCCCCTGGT GCTGATGGTC CTGCTGGTGC TCCCTGGTACT CCCCGGGCCTC AAGGTATTGC 2410 7420 7420 7420 7430 7440 7450 7460 ATCCCCTGGT GCTGATGGTC CTGCTGGTGC TCCCTGGTACT TCCCTGGTCT 24470 7480 7480 7480 7480 7480 7480 7480	1630	1640	1650	1660	1670	1680
1750 1760 1770 1780 1790 1890 AGCTGGTCTT CCAGGGCCTA AGGGTGACAG AGGTGATGCT GGTCCCAAAG GTGCTGATGG AGCTGGTCTT CCAGGGCCTA AGGGTGACAG AGGTGATGCT GGTCCCAAAG GTGCTGATGG CTCTCCTGGC AAAGATGGCG TCCGTGGTCT GACCGCCCC ATTGGTCCTC CTGGCCCTGC 1870 1880 1890 1900 1910 1910 1920 TGGTGCCCCT GGTGACAAGG GTGAAAGTGG TCCCACCGGC CCTGCTGGTC CCACTGGAGC 1930 1940 1950 1960 1970 1980 TCGTGGTGCC CCCGGAGACC GTGGTGAGCC TGGTCCCCCC GGCCCTGCTG GCTTTCCTGG 1990 2000 2010 2020 2030 2040 CCCCCCCTGGT GCTGACGCC AACCTGGTGC TAAAGGCGAA CCTGGTGATG CTGGTCCCAA 2050 2060 2070 2080 2090 2100 AGGCGATGCT GGTCCCCCTG GGCCTGCCG ACCCCTGGAC CCCCTTGGCC CCATTGGTAA 2110 2120 2130 2140 2150 2160 TGTTGGTGCT CCTGGAGCCA AAGGTGCTCG CGGCACAGGC GGCCCCCTG GTGCTACTGG 2170 2180 2190 2200 2210 2220 TTTCCCTGGT GCTGCTGGC GAAGGAGGC TCCTGGGCCCCCCCCCC	1690	1700	1710	1720	1730	1740
1810						
1870	AGCTGGTCTT	CCAGGGCCTA	AGGGTGACAG	AGGTGATGCT	GGTCCCAAAG	CTCCTCATCC
TGGTGCCCCT GGTGACAAGG GTGAAAGTGG TCCCACCGGC CCTGCTGGTC CCACTGGAGC 1930 1940 1950 1960 1970 1980 TCGTGGTGCC CCCCGGAGACC GTGGTGAGCC TGGTCCCCCC GGCCCTGCTG GCTTTGCTGG 1990 2000 2010 2020 2030 2040 CCCCCCTGGT GCTGACGGCC AACCTGGTGC TAAAGGCGAA CCTGGTGATG CTGGTGCCCAA 2050 2060 2070 2080 2090 2100 AGGCGATGCT GGTCCCCCTG GGCCTGCCGG ACCCGCTGGA CCCCCTGGCC CCATTGGTAA 2110 2120 2130 2140 2150 2160 2160 TGTTGGTGCT CCTGGAGCCA AAGGTCGTCC CGGCAXXGGT GGTCCCCCTG GTGGACCCCC GTGGACCCCC TCTGGAAATG CTGGAAAATG CTGGAAAATG CTGGAAAAATG CTCGGTGGTG AAAAAGAAGG CGGCAAAAGAT CCCCGTGGTG AAAAGAAAGG CGGCCAAAAGAT CCCCGTGGTG AAAAGAAAGG CCCCGTGGTG AGACTGCCCC AGACTGCCCC AGACTGCCCC <td< td=""><td>1810 CTCTCCTGGC</td><td>1820 AAAGATGGCG</td><td>1830 TCCGTGGTCT</td><td>1840 GACCGXXCCCC</td><td>1850 ATTGGTCCTC</td><td>1860 CTGGCCCTGC</td></td<>	1810 CTCTCCTGGC	1820 AAAGATGGCG	1830 TCCGTGGTCT	1840 GACCGXXCCCC	1850 ATTGGTCCTC	1860 CTGGCCCTGC
TCGTGGTGCC CCCGGAGACC GTGGTGAGCC TGGTCCCCCC GGCCCTGCTG GCTTTTCCTGG 1990 2000 2010 2020 2030 2040 CCCCCCTGGT GCTGACGGCC AACCTGGTGC TAAAGGCGAA CCTGGTGATG CTGGTGCCAA 2050 2060 2070 2080 2090 2100 AGGCGATGCT GGTCCCCCTG GGCCTGCCGG ACCCGTTGGA CCCCCTGGCC CCATTGGTAA 2110 2120 2130 2140 2150 2160 TGTTGGTGCT CCTGGAGCCA AAGGTGCTCG CGGCAXCGCT GGTCCCCCTG GTGCTACTGG 2170 2180 2190 2200 2210 2220 TTTCCCTGGT GCTGCTGCCC GAGTCGGTCC TCCTGGAAAAGA CCCCGTGGTG AGACTGGCCC 2230 2240 2250 2260 2270 2280 TGGCCCTCCT GGTCCTGCTG GCAAAAGAAGG CCCCGTGGTG AGACTGCCCC 2290 2300 2310 2320 2330 2340 TGCTGGACGT	1870 TGGTGCCCCT	1880 GGTGACAAGG	1890 GTGAAAGTGG	1900 TCCCAGCGGC	1910 CCTGCTGGTC	1920 CCACTGGAGC
CCCCCCTGGT GCTGACGGCC AACCTGGTGC TAAAGGCGAA CCTGGTGATG CTGGTGCCAA 2050 2060 2070 2080 2090 2100 AGGCGATGCT GGTCCCCCTG GGCCTGCCGG ACCCGCTGGA CCCCCTGGCC CCATTGGTAA 2110 2120 2130 2140 2150 2160 TGTTGGTGCT CCTGGAGCCA AAGGTGCTCG CGGCA:CGCT GGTCCCCCTG GTGCTACTGG 2170 2180 2190 2200 2210 2220 TTTCCCTGGT GCTGCTGGCC GAGTCGGTCC TCCTGGCAAATG CTGGACCCCC 2230 2240 2250 2260 2270 2280 TGGCCCTCCT GGTCCTGCTG GCAAAGAAGG CGGCAAAGGT CCCCGTGGTG AGACTGGCCC 2290 2300 2310 2320 2330 2340 TGCTGGACGT CCTGGTGAAG TTGGTCCCCC TGGTCCCCT GGCCCTGCTG GCGAGAAAGG ATCCCCTGGT GCTGATGGTC CTGCTGGTGC TCCTGGTACT CCCGGGGCCTC AAGGTATTGC 2410 2420 2430 2440 2450 2450 2460 TGGACAGCCT GGTGTGGTCG GCCTGCTGG TCAGAGAGGA GAGAGAGGCT TCCCTGGTCT	1930 TCGTGGTGCC	1940 CCCGGAGACC	1950 GTGGTGAGCC	1960 TGGTCCCCCC	1970 GGCCCTGCTG	1980 GCTTTGCTGG
AGGCGATGCT GGTCCCCCTG GGCCTGCCGG ACCCGTGGA CCCCCTGGCC CCATTGGTAA 2110 2120 2130 2140 2150 2160 TGTTGGTGCT CCTGGAGCCA AAGGTGCTCG CGGCACCGCT GGTCCCCCTG GTGCTACTGG 2170 2180 2190 2200 2210 2220 TTTCCCTGGT GCTGCTGCCC GAGTCGGTCC TCCTGGCCCC TCTGGAAATG CTGGACCCCC 2230 2240 2250 2260 2270 2280 TGGCCCTCCT GGTCCTGCTG GCAAAGAAGG CGGCAAAGGT CCCCGTGGTG AGACTGGCCC 2290 2300 2310 2320 2330 2340 TGCTGGACGT CCTGGTGAAG TTGGTCCCCC TGGTCCCCCT GGCCCTGCTG GCGAGAAAGG 2350 2360 2370 2380 2390 2400 ATCCCCTGGT GCTGATGGTC CTGCTGGTGC TCCTGGTACT CCCGGGCCTC AAGGTATTGC 2410 2420 2430 2440 2450 2460 TGGACAGCGT GGTGTGGTCG GCCTGCCTGG TCAGACAGAGAGAGGA GAGAGAGGGCT TCCCTGGTCT	1990 CCCCCTGGT	2000 GCTGACGGCC	2010 AACCTGGTGC	2020 Tahaggcgaa	2030 CCTGGTGATG	2040 CTGGTGCCAA
2110 2120 2130 2140 2150 2160 TGTTGGTGCT CCTGGAGCCA AAGGTGCTCG CGGCA:GGCT GGTCCCCCTG GTGCTACTGG 2170 2180 2190 2200 2210 2220 TTTCCCTGGT GCTGCTGCCC GAGTCGGTCC TCCTGGCCCC TCTGGAAATG CTGGACCCCC 2230 2240 2250 2260 2270 2280 TGGCCCTCCT GGTCCTGCTG GCAAAGAAGG CGGCAAAGGT CCCCGTGGTG AGACTGGCCC 2290 2300 2310 2320 2330 2340 TGCTGGACGT CCTGGTGAAG TTGGTCCCCC TGGTCCCCCT GGCCCCTGCTG GCGAGAAAGG 2350 2360 2370 2380 2390 2400 ATCCCCTGGT GCTGATGGTC CTGCTGGTGC TCCTGGTACT CCCGGGCCTC AAGGTATTGC 2410 2420 2430 2440 2450 2460 TGGACAGCGT GGTGTGGTCG GCCTGCCTGG TCAGAGAGGGA GAGAGAGGCT TCCCTGGTCT	2050 AGGCGATGCT	2060 GGTCCCCCTG	2070 GGCCTGCCGG	2080 ACCCGITGGA	2090 CCCCCTGGCC	Z100 CCATTGGTAA
2170 2180 2190 2200 2210 2220 TTTCCCTGGT GCTGCTGGCC GAGTCGGTCC TCCTGGCCC TCTGGAAATG CTGGACCCCC 2230 2240 2250 2260 2270 2280 TGGCCCTCCT GGTCCTGCTG GCAAAGAAGG CGGCAAAGGT CCCCGTGGTG AGACTGGCCC 2290 2300 2310 2320 2330 2340 TGCTGGACGT CCTGGTGAAG TTGGTCCCCC TGGTCCCCT GCCCCTGCTG GCGAGAAAGG 2350 2360 2370 2380 2390 2400 ATCCCCTGGT GCTGATGGTC CTGCTGGTGCT TCCTGGTACT CCCGGGCCTC AAGGTATTGC TGGACAGCGT GGTGTGGTCG GCCTGCCTGG TCAGAGAGGGA GAGAGAGGCT TCCCTGGTCT 2470 2480 2490 2500 7510 7510	2110	2120	2130	2140	2150	2160
2230 2240 2250 2260 2270 2280 TGGCCCTCCT GGTCCTGCTG.GCAAAGAAGG CGGCAAAGGT CCCCGTGGTG AGACTGGCCC 2290 2300 2310 2320 2330 2340 TGCTGGACGT CCTGGTGAAG TTGGTCCCCC TGGTCCCCT GGCCCTGCTG GCGAGAAAGG 2350 2360 2370 2380 2390 2400 ATCCCCTGGT GCTGATGGTC CTGCTGGTGCT TCCTGGTACT CCCGGGCCTC AAGGTATTGC TGGACAGCGT GGTGTGGTCG GCTGCCTGG TCAGAGAGAGA GAGAGAGGCT TCCCTGGTCT 2470 2480 2490 2500 7510 2620	2170	2180	2190	2200	2210	2220
2290 2300 2310 2320 2330 2340 TGCTGGACGT CCTGGTGAAG TTGGTCCCCC TGGTCCCCT GCCCCTGCTG GCGAGAAAGG 2350 2360 2370 2380 2390 2400 ATCCCCTGGT GCTGATGGTC CTGCTGGTGC TCCTGGTACT CCCGGGCCTC AAGGTATTGC 2410 2420 2430 2440 2450 2460 TGGACAGCGT GGTGTGGTCG GCCTGCCTGG TCAGAGAGAGA GAGAGAGGCT TCCCTGGTCT	. 2230	2240	2250	. 2260	. 2270	2280
2350 2360 2370 2380 2390 2400 ATCCCCTGGT GCTGATGGTC CTGCTGGTGC TCCTGGTACT CCCGGGCCTC AAGGTATTGC 2410 2420 2430 2440 2450 2460 TGGACAGCGT GGTGTGGTCG GCCTGCCTGG TCAGAGAGAGGA GAGAGAGGCT TCCCTGGTCT 2470 2480 2490 2500 7510	2290	2300	2310	2320	2330	2340
ATCCCCTGGT GCTGATGGTC CTGCTGGTGC TCCTGGTACT CCCGGGCCTC AAGGTATTGC 2410 2420 2430 2440 2450 2460 TGGACAGCGT GGTGTGGTCG GCCTGCCTGG TCAGAGAGAGGA GAGAGAGGCT TCCCTGGTCT 2470 2480 2490 2500 7510	TGCTGGACGT	CCTGGTGAAG	TTGGTCCCCC	TGGTCCCCCT	GGCCCTGCTG	GCGAGAAAGG
2470 2480 2490 2500 7510		2360 GCTGATGGTC	2370 CTGCTGGTGC	2380 TCCTGGTACT	2390 CCCGGGCCTC	2400 AAGGTATTGC
2470 2480 2490 2500 7510 7510	2410 TGGACAGCGT	2420 GGTGTGGTCG	2430 GCCTGCCTGG	2440 TCAGAGAGGA	2450 GAGAGAGGCT	2460 TCCCTOSTCT
	2470	2480	2490	2500	7510	2522

FIG. 16B

		2550	2560	2570	2580
2530	2540	VCC3	CCCTCCACCC	CCTCGTGAAT	CTGGACGTGA
TCCCCCCGGT	CCCATGGGCC	CCCCTGGATT	COC LOGUECE		
			2620	2630	2640
2590	2600	2010	2020	TOTAL CONCECTO	CCAAGGGTGA
CCCCCCTCCT	GCTGCCGAAG	CLICCCCIOC	YCOYOYCOO!		
				2400	2700 GTGCCCCTGG
2650	2660	2670	2680	oceconomic officer	CTCCCCCTGG
2650 CCGTGGTGAG	ACCGGCCCCG	CTGGACCCCC	Tecter.	COLOC ICE TO	0100000100
				2250	2760
2710	2720	2730	2740	2/30	2760
CCCCGTTGGC	CCTGCTGGCA	AGAGTGGTGA	TCGTGGTGAG	ACTGGTCCTG	CTGGTCCCGC
					2020
2770	2780	2790	2800	2810	2820
COGTCCCGTC	CCCCCCCCCC	GCGCCCGTGG	CCCCGCCGGA	CCCCAAGGCC	CCCGTGGTGA
00100010					2000
2830	- 2840	2850	2860	2870	2880
CAACCGTGAG	ACAGGCGAAC	AGGGCGACAG	AGGCAFAAAG	GGTCACCGTG	GCTTCTCTGG
G0.0001 G.0					
2890	2900	2910	2920	2930	2940
CCTCCAGGGT	CCCCTGGCC	CTCCTGGCTC	TCCTGGTGAA	CAAGGTCCCT	CTGGAGCCTC
2950	. 2960	2970	2980	2990	3000
ACCACALCA TA	CCTCCCCGAG	GTCCCCCTGG	CTCTGCTGGT	GCTCCTGGCA	AAGATGGACT
3010	3020	3030	3040	3050	3060
Carcinos	CORRECCE	TTGGGCCCCC	TGGTCCTCGC	GGTCGCACTG	GTGATGCTGG
				•	
3020	3080	3090	3100	3110	3120
meleneration.		רזירדינינארר	ACCAG: ACCC	CCTGGTCCTC	CCAGCGCTGG
3130	3140	3150	3160	3170	3180
TTTCGACTTC	ACCUTYCTYC	CCCAGCCACC	TCAAGLGAAG	GCTCACGATG	GTGGCCGCTA
				••••	
3190	3200	3210	3220	3230	3240
CTACCCCCCT	agatictCCC	TGGACACCAA	CUNTAL ALL	AGCTCCACGG	AGAAGAACTG
01	agarceocco	200.000			
3250	3260	3270	3280	3290	3300
CIRCCARCEC	Caccacacaca	TYCACTICCG	CAAGGACCTC	CCCTCCAACT	GGATCCACGA
c10c010c00	CHOCIGIACA	1104011000	crossocc	000100101	OWITCOICOI
3310	3320	3330	3340	3350	3360
CCCCARGGC	TACCATGCCA	»CTTCTCTC	CCCCCCCCC	こここ	GGAĢCCTGGA
occessor	MCCMIGCCA	ACTICIOCCI		CCCIACATII	ooviccioov
3370	3380	3300	3400	3/10	3420
CACCCACTAC	2000	WITCH CONTRACTOR	CD DCCI GCDM	3110	CCTCGGCGGC
CHECKOTAC	NOCANOUICC	1000001017	CONCOURT	MACCEGGGG	CCICOCCOC
3430	3440	3450	3460	2470	3480
2420	2440	7470	2400	07.00	740U
OCCUTOC TOC	GIGCCGCAGG	LOCHOOCE	GCTGCCCATC	GIGIACIACG	TGGGCCGCAA
3400	3500	2617	7574	3630	3540
2430	0000	CC22C2222	2240	1000 KACO	GCTGAtctag
occorroo16	OMPCMOC101.	CCANCATUAT	COLOCOCICE	1 GCANGIGCA	OCIGATORE
3550	1660	3570	3500	3500	3600
•					

FIG. 16C

10 QLSYGYDEKS	20 TGCISVPCPH	30 Gpsgprglpg	40 PPGAPGPQGF	. 50 QGPPGEPGEP	60 Gasgpmgprg
70 PPGPPGKNGD	80 DGEAGKPGRP	90 Gergppgp	100 ARGLPGTAGL	110 PGMKGHRGFS	120 GLDGAKGDAG
120	140 PGENGAPGQH	150	160	170	180
100	200 KGEAGPQGPR	210	220	230	240
250	260 PGARGPSGPQ	270	280	290	300
310		330	340	350	360
370	380	390	400	410	
430		450	460	470	480
490	500	510	520	530	540
550	PGEAGXPGEQ 560	570	580	590	600
APCVDGAXGD	AGAPGAPGSQ 620			•	660
VRGLTGPIGP	PGPAGAPGDK 680	GESGPSGPAG	PTGARGAPGD	RGEPGPPGPA	GFAGPPGADG
QPGAXGEPGD	AGAKGDAGPP	GPAGPAGPPG	PIĠNV 3APGA	KGARGSAGPP	GATGFPGAAG
730 RVGPPGPS@N	740 AGPPGPPGPA	GKEGGKGPRG	ETGPAGRPGE	VGPPGPPGPA	GEKGSPGADG
790 PAGAPGTPGP	QCIYCOYCA AASOYISO A	810 GLPGQRGERG	820 FPGLPGPSGE	830 PGKQGPSGAS	840 Gergppgphg
	660 SGREGAPAAE				
910 KSGDRGETGP	920. Agpagpvgpā	930 GargpagpQG	940 Pagdkgetge	950 QGDRGIKGHR	960 GFSGLQGPPG
		990 GPPGSAGAPG			1020 GDAGPVGPPG
1030 PPGPPGPPGP			106b GGRYYRARSD		1080 DDADFEPSLG
	1100 LRVVQCSDLG				1140 NYLLYNLYALI
					1200 ITKVRKVTFN
1210 GUYQHIVIEL		1230 Engapogiak			1260 SLTELYLDAN

FIG. 17A

1270	1280	1290	1300	. 1310	1320	
KISRVDAASL	KGLNNLAKLG	LSFNSISAVD	NGSLANTPHL	RELHLINNKL	TRVPGGLAEH	
					•	
1330	1340	1350	1360	1370	1380	
KYIQVVYLXN					STERCVYVES	
				gb.g.	0	
1390	1400	1410	1420	1430	1440	
AIOLGNYK*.						

10 QLSYGYDEKS	20 TGGISVPGPM	30 GPSGPRGLPG	40 PPGAPGPQGF	, 50 QGPPGEPGEP	60 Gasgphgprg
70 PPGPPGKNGD	80 DCEAGKPGRP	90 GERGPPGPQG	100 Arglpgtagl	. 110 PGHKGHRGFS	120 GLDGAKGDAG
130 PAGPKGEPGS	140 PGENGAPGOM	150 GPRGLPGERG	160 RPGAPGPAGA	170 RGNDGATGAA	180 GPPGPTGPAG
PPGFPGAVGA	200 KGEAGPQGPR	GSEGPQGVRG	EPGPPGPAGA	AGPAGNPGAD	GQPGAKGANG
250 APGIAGAPGF	260 PGARGPSGPQ	GPGGPPGPKG	NSGEPGAPGS	KGDTGAKGEP	GPVGVQGPPG
	RGEPGPTGLP		SRGFPGADGV	AGPKGPAGER	GSPGPAGPXG
370 SPGEAGRPGE	380 AGLPGAKGLT	390 GSPGSPGPDG	400 KTGPPGPAGQ	410 DGRPGPPGPP	420 GARGQAGVMG
430 FPGPKGAAGE	440 PGKAGERGVP	450 GPPGAVGPAG	460 XDGEAGAQGP	470 PGPAGPAGZR	480 GEQGPAGSPG
490 FQGLPGPAGP	PGEAGKPGEQ		PSGARGERGF	PĞERGVQGPP	GPAGPRGANG
550 APGNDGAKGD	560 AGAPGAPGSQ	570 GAPGLQGMPG	580 ERGAAGLPGP	590 KGDRGDAGPK	CYDCSECKDC 600
610 VRGLTGPIGP	620 PGPAGAPGDK	630 GESGPSGPAG	640 PTGARGAPGD	650 RGEPGPPGPA	660 GFAGPPGADG
670 QPGAKGEPGD		GPAGPAGPPG	PICNVGAPCA	KGARGSAGPP	Catcfpgaag
730 RVGPPGPSGN	740 AGPPGPPGPA	750 GKEGGKGPRG	760 ETGPAGRPGE	770 VGPPGPPGPA	780 GEKGSPGADG
790 PAGAPGTPGP		810 GLPGQRGERG			
PPGLAGPPGE		GSPGRDGSPG	AKGDRGETGP	AGPPGAXGAX	GAPGPVGPAG
910 KSGDRGETGP		930 GargpagpQG	PRGDKGETGE	950 QGDRGIKGHR	960 GFSGLQGPPG
970 PPGSPGEQGP	960 SGASGPAGPR				
1030 PPGPPGPPGP	1040 PSAGFDFSFL	1050 PQPPQEKAHD	1060 GGRYYRARSP	1070 KDLPPDTTLL	1080 DLQ:WKITEI
1090 KDGDFKVLKN	1100 LHALILVNNK				

FIG. 18

9 18 27 26 45 54 CAG CIG TOT THAT GGC THAT GAT GAG ANA TON ACC CGA GGA ATT TOO GTG COT GGC CCC ATG GOT CCC TOT GOT CCT CGT GOT CTC CCT GOC CCC CCT GGT GCA CCT GGT CCC CAA GGC TTC 139 147 ' 156 CHA GGT CCC CCT GGT GAG CCT GGC GAG CCT GGA GCT TCA GGT CCC ATG GGT CCC CGA GGT CCC CCA GGT CCC CCT CGA ANG ANT GGA GAT GAT GGG GAA GCT GGA ANA CCT GGT CGT CCT GGT GAG, CGT GGG CCT CCT GXG CCT CAG GGT GCT CCA GGA TTG CCC GGA ACA GCT GGC CTC CCT GGA ANG NAG GGA CAC AGA GGT TTC AGT GGT TTG GAT GGT GCC AAG GGA GAT GCT GGT CCT GCT GGT CCT ANG GCT GAG CCT GGC AGC CCT GGT GAA AAT GGA GCT CCT GGT CAG ATG GGC CCC CGT AGC CTG CCT GGT GAG AGA GGT CGC CCT GGA GCC CCT GGC CCT GGT GGT GCT COT GGA AAT GAT GOT ACT GOT GOT GCC GCG CCC CCT GGT CCC ACC GGC CCC GCT GGT CCT CCT GGC TIC CCT GGT GCT GTT GGT GCT AAG GGT GAA GCT GGT CCC CAA GGG CCC CGA GGC TOT GAA GGT CCC CAG GGT GTG CCT GGT GAG CCT GGC CCC CCT GGC CCT GCT GGT GCT CCT GGC CCT GCT GGA AAC CCT GGT GCT GAT GGA CAG CCT GGT GCT AAA GGT GCC AAT GGT GCT CCT GCT ATT GCT CGT GCT CCT GGC TCC CGT GCC CGA GGC CCC TCT GGA CCC CAG UGC CCC GGC GGC CCT CCT GGT CCC ANG GGT ANC AGG GGT GAN CCT GGT GCT CCT CGC AGG AAA OGA GAC ACT GGT CCT NAG OGA GAG CCT GGC CCT GTT GGT GTT CAA GGA CCC CCT GGC CCT GCT GGA GAG CAA GGA AAG CGA GGA GCT CGA GGT GAA CCC GGA CCC ACT GGC CTG CCC SCA CCC CCT GGC GAG CGT GGT CGA CCT GGT AGC CGT GGT TTC CCT GGC GCA GAT GGT GTT GET GET CCC AAG COT CCC GET GOT GAA CGT GGT TET CET GGC CCC GET GGC CCC AAA GGA TOT COT OUT GAA GOT COT COT COT GAA GOT GGT CTG COT GGT GCC AAG GGT CTG ACT CON MOST COT GGC AGC CUT GGT CCT GAT GGC ANN ACT GGC CCC CCT GGT CCC GGT CAN

FIG. 19A

1269 1278 1287 1296 THE COT GRAINST MAN CAT GOT GOT GRAINGS COD GGS AND GOT GRAING GOT GOT GOT GOT GOT GGA CCC COT CUC GCT GTC GGT CCT GCT GCC AAA GAT GGA GAG GCT GGA GCT CAG GGA CCC CCT GGC CCT GCT GGT CCC GCT GGC GAG AGA GGT GAA CAA GGC CCT GCT GGC TCC CCC GGA TTC CAG GGT CTC CCT GGT CCT GGT GGT CCT CCA GGT GAA GCA GGC AAA CCT GGT GAA CAG GGT GTT CCT GGA GAC CTT GGC GCC CCT GGC CCC TCT GGA GCA AGA GGC GAG AGA GGT TTC CCT GGC GAC CGT GGT GTG CAA GGT CCC CCT GGT CCT GGA CCC CGA GGG GCC AAC GGT GCT CCC GCC AAC GAT GCT GCT AAG CGT GAT GCT GGT GCC CCT GGA GCT CCC GGT AGC CAG GGC GCC CCT GGC CTT CAG GGA ATG CCT GGT GAA CGT GGT GCA GCT GGT CTT CCA GGG CCT .1758 ARG GGT GAC AGA GGT GAT GCT CGT CCC AAA GGT CCT GAT GGC TCT CCT GGC AAA GAT GGC 1845 1854 GYC CGT GGT CTG ACC GGC CCC ATT GGT CCT CCT GGC CCT GCT GGT GCC CCT GGT GAC AAG GGT GAA AGT GGT CCC AGC GGC CCT GGT GGT CCC ACT GGA GCT CGT GGT GCC CCC GGA GAC CET CET GAS CCT GET CCC CCC GGC CCT GCT GGC TIT GCT GGC CCC CCT GGT GCT GAC GGC CAA CCT CCT GCT AAA GGC GAA CCT GGT GAT CCT GGT GCC AAA GGC GAT GCT GGT CCC CCT COG CCT GCC GGA CCC GCT GGA CCC CCT GGC CCC ATT GGT AAT GTT GGT GCT CCT GGA GCC ANA OUT GOT COC COC AGO COT GOT CCC COT GOT GCT ACT GOT TTC CCT GGT GCT GCT GGC CGA GTC GGT CCT CGT CGC CCC TCT GGA AAT GCT GGA CCC CCT GGC CCT CCT GGT CCT GCT 2238 · GGC AAA GAA GGC CGC AAA GGT CCC CGT GGT GAG ACT GGC CCT GCT GGA CGT CCT GGT GAA GTT GGT CCC OUT GGT CCC CCT GGC CCT GGT GGC GAG AAA CGA TCC CCT GGT GCT GAT GGT CCT OCT GOT CCT CCT GOT ACT CCC COG CCT CAA GOT ATT GCT GGA CAG CGT GGT GTG GTC GGC CTC CCT CAT CAG AGA GGA GAG AGA GGC TTC CCT GGT CTT CCT GGC CCC TCT GGT GAA COT COC ARA CAR GOT CCC TOT GGA GCA AGT GGT GAR CGT GGT CCC CCC GGT CCC ATG GGC

FIG. 19B

2529 2538 2547 2556 2.05 2574 CCC CCT GGA 11G CCT GGA CCC CCT GGP GAA TCT GGA CGT GAG GGG GCT CCT GCT GCC GAA GGT TCC CCT GGA CGA GAC GGT TCT CCT GGC GCC AAG GGT GAC CGT GGT GAG ACC GGC CCC AND ACT COT CAT COT CAT CAS ACT CCT CCT CCT GCT CCC GCC GCT CCC GTC CGC CCC 2757 2796 GCC GCC CGT GCC CCC GCC GGA CCC CAA GGC CCC CGT GGT GAC AAG GGT GAG ACA GGC GAA CAG GGC GAC AGA GGC ATA AAG GGT CAC CGT GGC TTC TCT GGC CTC CAG GGT CCC CCT GGC CUT COT GGC TOT COT GGA CAA CGT CCC TOT GGA GCC TOT GGT CCT GGT CCC CGA GGT CCC CCT GGC TCT QCT GGT GCT CCT GGC AAA GAT GGA CTC AAC GGT CTC CCT GGC CCC ATT GGG CCC CCT GGT CCT CGC GGT CGC ACT GGT GAT GCT GGT CCT GTT GGT CCC CCC GGC CCT COT GGA CCT CCT GGT CCC CCT GGT CCT CCC AGC GCT GGT TTC GAC TTC AGC TTC CTC CCC CAG CCA CCT CAA GAG AAG GCT CAC GAT GGT GGC CGC TAC TAC CGG GCT AGA TCC GAT GAG GOT TOT GGG ATA GCC COA GAA GIT COT GAT GAC CGC GAC TTC GAG CCC TCC CTA GGC CCA GING TOO CCC THE COC TOT CAR TOO CAT CIT COR GIG GIC CAG TOT TOT GAT TIG GOT CTG GAC AAA GTG CCA AAG GAT CTT CCC CCT GAC ACA ACT CTG CTA GAC CTG CAA AAC AAC ALA ATA ACC GAA ATC AAA GAT GGA GAC TIT AAG AAC CTG AAG AAC CTT CAC GCA TIG ATT CTT GTC AAC AAT AAA ATT AGC AAA GIT AGT CCT GGA GCA TTT ACA CCT TIG GTG AAG TTG GAA CGA CTT TAT CTG TCC AAG ANT CAG CTG AAG GAA TTG CCA GAA AAA ATG CCC AAA ACT CIT CAG GAG CTG CCT GCC CAT GAG AAT GAG ATC ACC AAA GTG CGA AAA GTT ACT TTC AAT GGA CTG AAC CAG ATG ATT GTC ATA GAA CTG GGC ACC AAT CCG CTG AAG AGC TCA GGA ATT GAN ANY GOO OCT THE CAS GON AND AND CHE TEE THE ATE COE ATT GET GAT ACE ANT ATC ACC AGC ATT CCT CAA GGT CTT CCT TCC CTT ACG GAA TTA CAT CTT GAT GGC AAC

FIG. 19C

AMA	3789 ATC -AGC	AGA	3798 GTT CAT	GCA	3907 GCT AGC	cīc	3916 AAA GGA	CTG	3825 AAT AAT	TIG	3834 CCT AAG	TTG GGA
ידויG	3849 AGT TTC	NVC.	3858 AGC ATC	TUT	3867 GCT GTT	GAC	3876 CGC TAĄ	TCT	3885 CTG GCC	AAC	3894 ACG CCT	CAT CTG
	3909		3918		3927		3936		3945 CCT GGT		3954	
	3969		3978		3997		3996		4005 TCT GTA		4014	
	4029		4038		4047		4056		4065 TCG GGT		4074	
	4089		4098		4107		4116		4125 AGA TGT		4134	
	4149		4158		4167		4176		4185		4194	

FIG. 19D

. 10 . gggaaggatt	20 tccatttccC	30. AGCTGTCTTA	40 TGGCTATGAT	50 GAGAAATCAA	06 TAADDADDOO
	80 DOCCCCATGG	90	100	110	120
			1.00	120	180
ACCTGGTCCC	CAAGGCTTCC	AXGGTCCCCC	220	230	240
TCCCATGGGT	CCCCGAGGTC	CCCCACCICC	CCCIOONA	,0,100.10	
TGGAAAACCT	260 GGTCGTCCTG	C.I.CACCC100	OCC.CC.1000	00.0	
310 GCCCGGAACA	320 GCTGGCCTCC	066 A÷DTAKADTÇ	340 GGGACACAGA	350 GGTTTCAGTG	360 GTTTGGATGG
		200	400	410	420 CTGGTGAAAA
		450	450	470	480
TGGAGCTCCT	' GGTCAGATGG	CCCCCCTGG	CCIGCCIGGI	CAGAGAGAIC	GCCCTGGAGC 540
CCCTGGCCCT	CTGGTGCTC	GTGGAAATGA	TGGTGCTACT	. GGIGCIGCCG	GGCCCCIO
550 TCCCACCGGC	560 CCCGCTGGTC	570 CTCCTGGCTT	580 CCCTGGTGCT	590 CTTGGTGCTA	000 AGGGTGAAGC
610 TGGTCCCCAX	620 GGGCCCGAG	630 GCTCTGAAGG	640 TCCCCAGGGT	650 GTGCGTGGTG	660 ACCTGCCC
670	680	690	700	710	720 GACAGCCTGG
730	740	750	760	770	780
					CTGGTGCCCG 840
AGGCCCCTCI	GGACCCCAGG	cccccccc	CCCTCCTGGT	CCCAAGGGTA	ACAGCGGTGA
850 ACCTGGTGCT	860 CCTGGCAGCA	870 AAGGAGACAC	1GGTGCTAAG	890 GGAGAGCCTG	900 GCCCTGTTGG
910 TGTTCAAGGA	920 CCCCCTGGCC	930 CTGCTGGAGA	940 Dakaddaada	950 CGAGGAGCTC	960 GAGGTGAACC
970 CGGACCCACT	980	990 GACCCCTGG	1000 CGAGCGTGGT	1010 GGACCTGGTA	1020 GCCGTGGTTT
1030	1040	1050	1060	1070	1080 GTTCTCCTGG
		1110			
CCCCCCTGGC	CCCAAAGGAT	CTCCTGGTGA	AGCTGGTCGT	CCCGGTGAAG	CTGGTCTGCC
1150 TGGTGCCAAG	1160 CGTCTGACTG	1170 GAAGCCCTGG	1180 CAGCCCTGGT	1190 CCTGATGGCA	1200 AAACTCGCCC
1210 CCCTCGTCCC	1220 CCCGGTCAAG	1230 ATGGTCGCCC	1240 CGGACCCCCA	1250 GGCCCACCTG	1260 GTGCCCGTGG

FIG. 20A

	•		-		
1270 TCAGGCTGGT	1280 GTGATGGGAT	1290 TCCCTGGACC	1300 TAAAGGTGCT	. 1310 GCTGGAGAGC	1320 CCGGCAAGGC
1330	1340	1350	1360	1370	1380 AAGATGGAGA
1390	1400	1410	1420	1430	1440
GGCTGGAGCT	CAGGGACCCC	CTGGCCCTGC	TECTCCCECT	GGCGAGAGAG	GTGAACAAGG
1450 CCCTGCTGGC					1500 CAGGTGAAGC
1510 AGGCAAACCT	1520 GGTGAACAGG	1530 GTGTTCCTGG	1540 AGACCTTGGC	1550 GCCCCTGGCC	1560 CCTCTGGAGC
1570 AAGAGGCGAG	1580 AGAGGTTTCC	1590 CTGGCGAGCG	1600 TGGTGTGCAA	1610 GGTCCCCCTG	1620 GTCCTGCTGG
1630		1650	1660	1670	1680
	1700				
TGGAGCTCCC	GGTAGCCAGG	CCCCCCTGG	CCTTCAGGGA	ATGCCTGGTG	AACGTGGTGC
1750 AGCTGGTCTT	1760 CCAGGGCCTA	1770 AGGGTGACAG	1780 AGGTGATGCT	1790 GGTCCCAAAG	1800 GTGCTGATGG
1810 CTCTCCTGGC	1820 AAAGATGGCG	TCCGTGGTCT	1840 GACCGGCCCC	1850 ATTGGTCCTC	1860 CTGGCCCTGC
1870 TGGTGCCCCT	1880 GGTGACAAGG	1890 GTGAAAGTGG	1900 TCCCAGCGGC	. 1910 CCTGCTGGTC	1920 CCACTGGAGC
1930	1940	1950	1960	1970	1980
	CCCGGAGACC	GTGGTGAGCC	TOGTCCCCCC	GCCCTGCTG	GCTTTGCTGG
CCCCCCTGGT	2000 GCTGACGGCC	2010 AACCTGGTGC	2020 TAAAGGCGAA	2030 CCTGGTGATG	2040 CTGGTGCCAA
2050 AGGCGATGCT	2060 GGTCCCCCTG	2070 GGCCTGCCGG	2080 ACCCGCTGGA	2090 CCCCCTGGCC	2100 CCATTGGTAA
	2120	2130	2140	2150	21.60
	2180				
TTTCCCTGGT	CTCTCTCCC	CAGTCCGTCC	TCCTGGCCCC	TCTGGAAATG	CTGGACCCCC
TGGCCCTCCT	2240 CGTCÇTCCTG	2250 GCAAAGAAGG	2260 CGGCAAAGGT	2270 CCCCGTGGTG	2280 AGACTGGCCC
	2300	2310	2320	2330	2340
	2360	2370	2380	2390	2400
2410	2420	2430	2440	2450	2460
TGGACAGCGT	GGTGTGGTCG	GCCTGCCTGG	TCAGAGAGGA	GAGAGAGGCT	TCCCTCGTCT
2470 TCCTCGCCCC	2480 TCTGGTGAAC	2490 CTGGCAAACA	2500 AGGTCCCTCT	2510 GGAGCAAGTG	2520 GTGAACGTGG

FIG. 20B

-				•	
2530	2540	2550	2560	. 2570	2580
					CTGGACGTGA
2590	2600	2610	2620	2630	2640
GGGGGCTCCT	GCTGCCGAAG	CTTCCCCTGG	ACGAGACGGT	TCTCCTGGCG	CCAAGGGTGA
2650	2660	2670	2680	2690	2700
CCGTGGTGAG	ACCGGCCCCG	CTGGACCCCC	TGGTGCTCNT	GGTGCTCNTG	GTGCCCCTGG
2710	2720	2730	2740	2750	2760
				ACTGGTCCTG	
		•			
2770	2780	2790	2800	2810	2820
CGGTCCCGTC	GGCCCCGCTG	CCCCCCCTGG	CCCCGCCGGA	CCCCAAGGCC	CCCGTGGTGA
	•				
2830	2840	2850	2860	2870	2880
CAAGGGTGAG	ACAGGCGAAC	AGGGCGACAG	AGGCATAAAG	GGTCACCGTG	GCTTCTCTGG
2800	2900	3010	2020	2930	2242
CCTCCACCCT	2300	2310	2920	2930	2940
				CAAGGTCCCT	
2950	2960	2970	2980	2990	3000
TEGICCIECT	GGTCCCCGAG	GTCCCCCTGG	CACACCACCA	GCTCCTGGCA	000C
3010	3020	3030	3040	3050	3060
CAACGGTCTC	CCTGGCCCCA	TTGGGCCCCC	TGGTCCTCGC	GGTCGCACTG	COCCARCOLO
3070	3080	3090	3100	3110	3120
1CC1G11GGT	CCCCCCCCCCCC	CTCCTGGACC	TCCTGGTCCC	CCTGGTCCTC	CCACCCCTCC
3130	3340				-2.0000.00
TTTCGACTTC	3140	3120	• 3160	3170	3180
	AGCITCCTCC	CCCAGCCACC	TCAAGAGAAG	3170 GCTCACGATG	GTGGCCGCTA
3190	3200	2210	•	3230	
CTACCGGGCT	ACET CECCAA	3210	3220	3230	3240
			CCCIWICACA	ACICIGCTAG .	ACCTGCAAAA
3250	3260	2020			
CAACAAAATA	ACCGAAATCA	U SC ADGDOVEDEA	3280	3290	3300
		and the order	CITIAAGAAC	CIGAAGAACC '	TTCACGCATT
3310	3320	2220			
GATICITOIC,	AACAATAAAA 1	TTAGCAAAGT		3350	3360 ·
			runter TORY	raactgcag.	

FIG. 20C

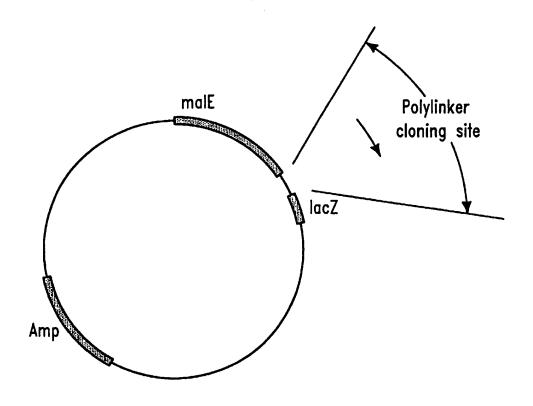
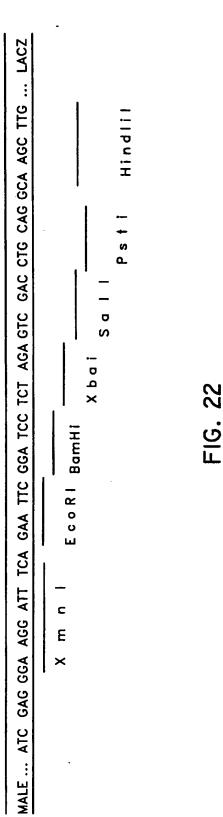


FIG. 21



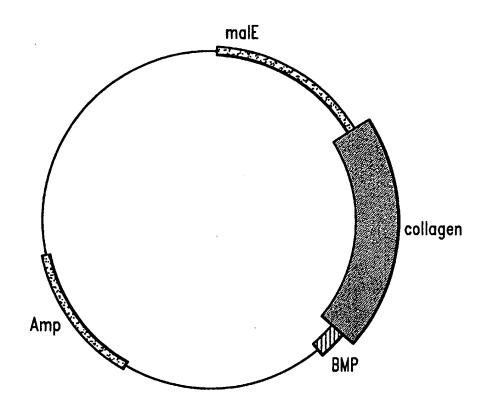


FIG. 23

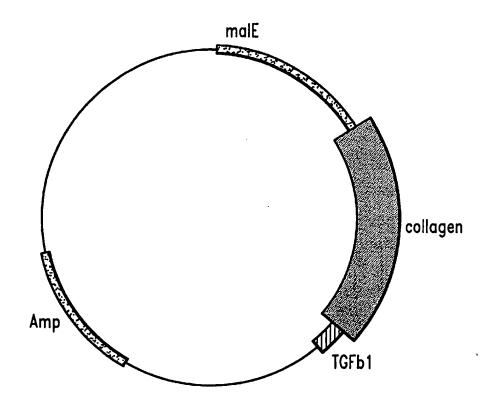


FIG. 24

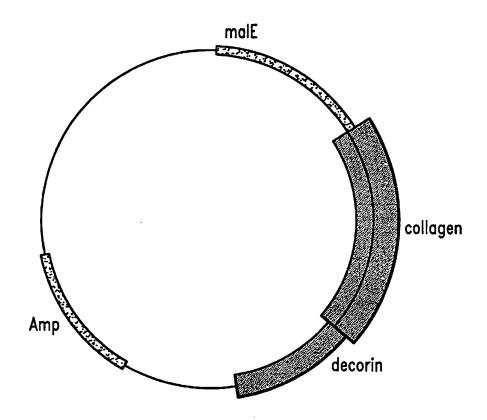


FIG. 25

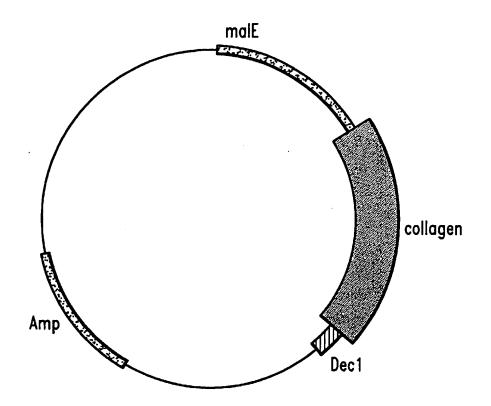


FIG. 26

			0			18		٠	27			36			45			54
5.	ಘ್ರ	CTG	TCT	TAT	GGC	TAT	Cat	GAG	AAA	TCA	ACC	GGA	CCY	Y.	700	GTG	cc:	
•								G1::	Tue	Ser	Thr	Glv	GLv	īle	Ses	Val	Pro	Gly
	Gin	Leu	Ser	Tyr	Gly	Tys	wzń	GIG	Lys									108
			63			72			81	CTC	ررية	90	ccc	œ.	25. 68	ecr	CCT	
	CCC	ATG	CGT	ccc	TCT	GGT	CCT	CGI	COL									GGT
	2-7	Mes	Glv	220	Ses	Gly	Pro	Arg	Gly	Leu	Pro	GŢĀ	Pro	5:0	GŢĀ	Ala	210	GŢÅ
									125			144			153			152
		רא	117	TTC	CAA	GGT	ccc	ÇCT	<u> </u>	GAG	CCT	GGC	G-AG	α :	CCH	GCT	TCA	GGT
	320	Gln	Gly	Phe	Gin	GTÀ	550	520	GT.	Giu	510	Ģīy	020		;			GŢĀ
			171			180			189			198		***	207	-,-	C17	216
																		GGG
	220	Met	Gly	520	Arg	Gly	Pro	Pro	Gly	Pro	210	Gly	Lys	Asn	Gly	775	λs	Gly
						234			243			252			251			270
	SAA	GCT	225 GCA	AAA	. œ:	GGI	CGI	CCI	GGT	G+G	CGT	GGG	∞	Œ			CYC	GGT
														4				~~~
	Glu	Ala	GLy	Lys	Pro	GTA	Arg	Pro	er.	GTR	وتم	GTĀ	. 510	810	GTŽ	720	GIL	GŢĀ
			279			288			297			306			315			324
	GCT	CC	. 633	TTG	: 000	GGA	AC	GCI	- GEV	CTC	<u> </u>	G	ATG	يهدر				GET
	AL a	Arg	Gly	Leu	250	GLy	The	: Ala	Gly	Leu	?ro	Gly	Mec	Lys	GLy	His	Arg	G1y
						342			351			360	1		369			378
	770	AGT	GGT GGT		GAT			AAG			CCI			œ			AAG	GT
										·								
	rne	261	GTĀ	ren	, ASD	GIY	MIC	. בעַנ	G.V	rsp		GTĀ	710	علم	<u> </u>		ېږي	GīĀ.
			387			396			405			414			423			432
	ريسين	CC	ست	بنجاء		GGT												GGC
	Glu	Pro	Gly	Ser	210	Gly	Glu	ลรถ	Gly	Ala	Fro	Gly	Gla	Met	Gly	Fro	وعذ	Gly
			441			450			459			468			477			455
	CTG	CCT		GEG	AGA.		೦೦೦	CCI		α	∞			GCT		GCT	Œ	CC'Y
				~~~														
	<b></b>	720	G.Y	GIU	44.5	GTĀ	λtų	210	σñ	W10	210	وين	220	ni:	ينى	عـــ	ar ç	Gly
		~~=	495			504			513	^~~		522			531			540
	AA1	GAT	GGT		AL:	GGT	GCT			<u></u>	ccr	GGT	ccc	ACC	Œ	<u> </u>	GUE	Gui
	۸s۳	Asp	Gly	Als	The	Gly	Ala	Ala	Gly	Pro	?rc	Gly	220	The	GL;	250	Ala	Gly
	•		549			558			567			576			553			594
	CC:	CCT	GGC	TTC	ಯ		GCT	GIT		GCT	AAG		GAA	cc:	œ	ಯ		GGG
	220	220	Gly	216	210	ĊΤĀ	ALZ	AST.	GTÅ	ATS	٦Å2	CTA	GLU	YIS	Giy	220	Gin	GiV
			603			612		<b></b> -	621			630			639			648
	ccc	CGA	CGC	TCT	GAA	GGT	ccc	CAC	GGT	ere.	CGI	GGT	GAG	œ:	GGC	ccc	CCT	GGC
	220	Arg	Gly	Ser	Glu	Gly	Pro	Gln	Gly	۷ <u>al</u>	λrç	Gly	Glu	?ro	GLy	220	220	GLy
		J					•				-							•
	cct	GCT	GGT	CCT:	GCT	666 GGC	CCI	GCT	675 CC:	AAC	CCT	684 GGT	GCT	GAT	693 GGA	CAG	CCT	702 GGT
	Pro	λĻε	GīÀ	715	<u>212</u>	Gī'n	510	λla	Gr.	٠.5.٦	?::	GŢÅ	λlε	çań	3.;	51:	250	Gly

FIG. 27A

GCT	Aaa	711 GGT	GCC	<b>A3</b> T	720 GGT	GCT	CCT	729 GGT	ATT	GCT	738 GGT	507	œ	747 GGC	: TTC	CCT	75 GG:
															Phe		
	-,-	765			774			783			792			801			810
GCC	CGY	GGC	$\alpha$	TCT	GGA	$\alpha$	CYC	œ	ccc	GGC	GGC	CCT	CCT	GGT	CCC	AAG	GGT
Ala	Arg	GlŸ	5.0	Şer	Gly	Pro	Gin	Gly	Pro	Gly	Gly	920	Sto	Gly	5ro	Lys	Gly
AAC	AGC	819 GGT		CCI	828 GGT				AGC				ACT	855 GGT	CT	aag	
, Asn	Ser	Gly	Glu	Pro	GŢĀ	Alz	5:0	Gly	Ser	Lys	Gly	 Ç2A	Thr	Gly	Ala	Lys	GJ À
616		873		c==			<b>C</b>				900		~~~	909		٥.,	918
															GAG Glu		
014	220	927		VG.	936	167	0211	945			954		***	963		GIU	972
AAG	CGA			CGA		GAA	$\frac{1}{2}$			ACT			œς 		<u>ccc</u>	CCT	GGC
Lys	Arg	GŢĀ	Ala	yzd	Gly	Glu	Pro	GŢĀ	510	The	Gly	Leu	920	Gly	Pro	Pro	Gly
GYG	ŒĨ	981 GGT		cc:	990 GGT	AGC	CGT	999 GGT	TIC	1 Œ!	.008 GGC	GCA	eat Eat	1017 GGT	GII	CCT	.026 GGT
							~~~								 Val		
ccc		.035	œ	رــــا ا	044 GGT :	623 .	رت ة 1	053	ىتىداق	1	062		1	071	ccc .	1	080
															 Pro :		
	1	230		1	98		1	107		1	116		,	125			
							<u> </u>	SET (GAA (CT (GGT (TTG (X7 (GGT	GCC 2	aac (SGT
Ser :			Glu i			fid i	Pro (Gly (Giu 2	₹Ţ\$ (ly:	leu :	?=> (Gly :	Ala I	iys (31 <u>y</u>
CTG 2		143 3GA 3	AGC (11 CCT (152 GC A	ec (CT (161 ST (ce c	11 AT 0	170 IGC 3	ea a	11 CT (179 SCC (CC C	li CT G	.88 :GT
Leu 1																	
	11	.97		12	06		12	15		12	24						
CCC C																	
Pro A																	
CAG G		GT G	TG A		GA T	rc o	CT G	ey c	CT A	93 G	78 GT G	CT G	CT G	Gy G	3G O	12 C. G	96 SC
Gln A	Js C	ly V	त्र स	et G	Ly Pl	ne P	ro G	ly P	ro L	vs G	ly A	la A	la G	ly G	lu P	ro G	ly
AAG G	13 CT G		AG CI	13: Ca G		rr co	13: C G	23 SA CO	$\infty \alpha$	13: T G	32 3C G	T G	13 13	41 ST C	רים יידי	139 T C	50
Lys A																	
													,		• •		-2

FIG. 27B

					1404
AAA GAT GGA GA	1368	1377	1386	CCT GCT GG7	222 722 222
AAA GAT OGA GA	G CCT CGA	GE CHG GEA			
Lys Asp Gly Gl	lu Ala Gly	Ala Gla Gly	9:0 Pro Gly	bio yra Già	550 YTS GTA
			1440	1449	1458
CAG AGA GGT G	ay cay eec	CCT GCT GGC	TOC CCC GGA	TTC CAG GGA	C:C CC1 GG1
Glu Arg Gly G	lu Gln Gly	Pro Ala Gly	Ser Pro Gly	Phe Gln Gly	Ten Sto CTA
		. 1405	1494	1503	1512
1467 CCT GCT GGT C	CT CCA GGT	ಲು ಜು ಜ	AAA CCT GGT	GAA CAG GGT	GIT CCT GGA
Pro Ala Gly P		Glu Aia Gly	· Lvs Pro Gly	Glu Gla Gly	Val Pro Gly
SEO YES GIÁ S					
1521 GAC CTT GGC G	1530 CCT GGC)	א פכא אפא פפט דיים	COR AGA GGT	
Asp Leu Gly A					
Asp Leu Gly A	Ta 520 CT	•			
1575 GAG CGT GGT G	1584 	1593 F CCC CCT GG	1602 TOTT GCT GGA	1611 CCC CGA GGG	1620 GCC AAC GGT
Glu Arg Gly V	ST GTV GT	y Pro Pro Gl	i suo yrg era	, 510 WLG CTA	KIS ASH GIY
1629	1638	164	1656	1665	1674
GCT CCC GCC A				~-	
Ala Pro Gly A	en Asp Gl	y Ala Lys Gi	y yeb yys Cyñ	Ala Pro Gly	· Ala Pro Gly
1683	1692	2 170			
Yes cyc eec e	CCT CCC	con cas co		GYY COLL COLL	GCY GCL GGL.
Ser Gim Gly A	la Pro Gly	/ Leu Gla Gly	Met Pro Gly	Glu Arg Gly	Ala Ala Gly
1737	1746	5 175	1764	1773	
CTT CCA GGG C	CT AAG GG1	. etc 167 etc	GAT GCT GGT	CCC AAA GGT	GCT GAT GGC
Leu ?ro Gly ?	zo Lys Gly	Asp Arg GL	y كاع والا	Pro Lys Gly	Ala Asp Gly
1791	1800	1809	1818	1827	1836
TOT TOT GGG A					
Ser Pro Gly L	vs Aso Glv	Val Are Gly	Leu Thr Glv	Pro Ile Glv	Pro Pro Gly
_			_	_	
1845 CCT GCT GGT G	1854 CC CCT GGT		1872 Gaa agt ggt		1890 CCT GCT GGT
Sto Ala Gly A	-		_	, ,	ALO MIN GIÀ
1899	1908	1917	1926	1935	1944
DCD ACT GGA G					
Pro Thr Gly A	la Arg Gly	Ala Pro Gly	yeb yed Cla	Glu Pro Gly	Pro Pro Gly
1953	1962			1989	1998
CCT GCT GGC T	TT GCT GGC	CCC CCT GGT	GCT GAC GGC	CAA CCT GGT	GCT AAA GGC
Pro Ala Gly P	he Ala Gly	Pro Pro Gly	Ala Asp Gly	Gln Pro Gly	Ala Lys Gly
2007	2016			2043	2052
GAA ÓCT GGT G	AT GCT GGT	CCC NAY CCC	GAT GCT GGT	CCC CCT GGG	CCC CCC CCA
Glu Pro Gly A	sp Ala Gly	Ala Lys Gly	Asp Ala Gly	Pro Pro Gly	Pro Ala Gly

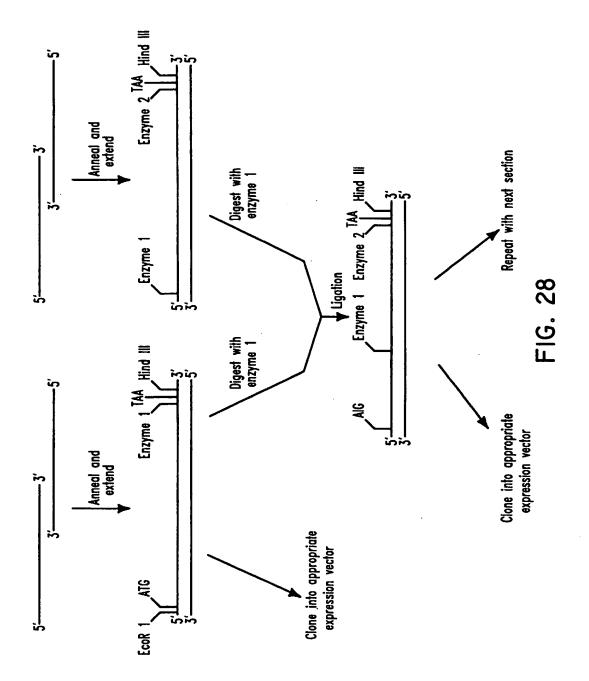
FIG. 27C

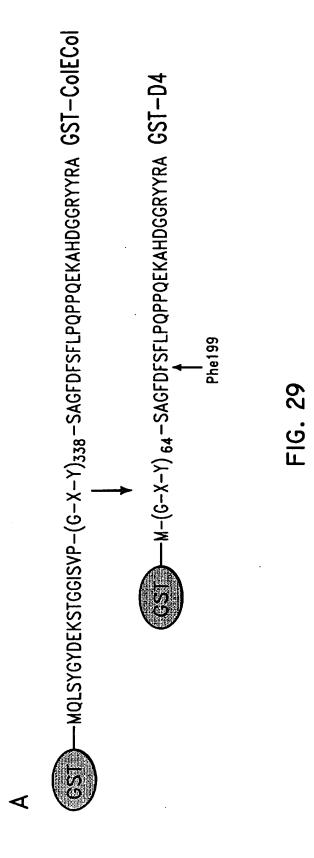
•													
2061	2070		2	079		:	2088			2097			2106
CCC CCI CCY CCC	CCT GGC	CCC	ATT	CCT	aat	GTT	GGT	GCT	CCT	GGA	GCC	AAA	
Pro Ala Gly Pro	Sto Glå	bio	ile	CIÃ	ASN	٧٨١	GIĀ	YTA	220	Orā	~~~	0,0	
2115	2124		2	133		:	2142			2151			2160
2115 GCT CGC GGC AGC	GCT GGT	CCC	CT	CCT	GCT	ACT	GGT	TTC	α T	GGT	GCT	CCT	GGC
Ala Arg Gly Ser	Ala Gly	Pro	520	CŢĀ	Ala	The	GĮλ	Phe	Pro	ĊŢĀ	Ala	γγa	GTÅ
,										2205			2214
2169 CGA GTC GGT CCT	2178		*****	2187	ስድጥ	COT	2196 663	CCC					
<i>व्हर वाट व्हा व्या</i>	CCT GGC												
Arg Val Gly Pro	Pro Glv	Pro	Ser	Gly	Asn	Ala	Gly	Pro	Pro	Gly	9ro	Pro	Gly
Arg var or,													
2223	2232		2	2241			2250			2259			2268
OCT GCT GGC ANA	GRA GGC	GGC	AAA	GGT	ccc	CGT	GGT	CAL:	ALT		···	GC 1	المتحا
Pro Ala Gly Lys	Clu Clu	Glu	T.US	Glv	Pro	7:0	Glu	Glu	Thr	Glv	Pro	Ala	Glv
ALC WIS GIA TAX	Gra G-y	U.,	-,,,	- -,		,	;						•
2277	2286	;	2	2295			2304			2313			2322
व्या व्या व्या	GTT GGT	∞	œ:	GG:	∞	CCT	GGC	CCI	CCT	GGC	GAG	AAA	GGA
Arg Pro Gly Glu	. Val Clu	D-0	3	Gly	P.00	250	Glu	Pro	Ala	Glv	Glu	Tus	GIV
Ard ato ora dra	. ver Gry	0		0-7			Ory		,			-,-	- -3
)					2358			2367			2376
TOC OCT GGT GCT	GAT GGT	CCI	CCT	ŒŢ	CCI	CCI	ĠŒĨ	ACT	∞	GGG	α	CYY	GGT
				~~~			~			Cl.,			<u></u>
Ser Pro Gly Ala	ASO GIV	550		G.A.A.	ALC.		(117	107	220	GLY	==0	Gitt	- G-V
-										4442			•
2385	· 2394		ž	2403		:	2412		٠.	2421			2430
2385 ATT GCT GGA CAG	· 2394		ž	2403		:	2412		٠.	2421			2430
ATT GCT GGA CAG	2394	GIG	GTC	2403 ESSC	CIG	CCT	2412 GGT	CAG	AGA	2421 GGA	CAG	AGA	2430 GGC
2385 ATT GCT GGA CAG Ile Ala Gly Gin	2394	GIG	GTC	2403 ESSC	CIG	CCT	2412 GGT	CAG	AGA	2421 GGA	CAG	AGA	2430 GGC
ATT CCT CGA CAG	2394	GIG Val	GTC  Vai	2403 ESSC	CIG	CCT  ?ro	2412 GGT	CAG GLn	AGA  Arg	2421 GGA	GAG Glu	AGA Arg	2430 GGC
ATT GCT GGA CAG	2394 CGT GGT Arg Gly	GIG Val	GTC  Val	2403 GGC Gly 2457	CTG  Leu	CCT  Pro	2412 GGT  Gly 2466	CAG GLn	AGA Arg	2421 GGA  Gly 2475	GAG Glu	AGA Arg	2430 GGC Gly 2484
ATT GCT GGA CAG  Ile Ala Gly Gla  2439  TTC CCT GGT CTT	2394 CGT GGT Arg Gly 2446	CCC Val	GTC Val	2403 660 Gly 2457 667	CTG Leu GAA	CCT	2412 GGT Gly 2466 GGC	CAG GLn AAA	AGA Arg CAA	2421 GGA Gly 2475 GGT	GAG Glu	ACA Arg	2430 GGC Gly Gly 2484 GGA
ATT GCT GGA CAG Ile Ala Gly Gin 2439	2394 CGT GGT Arg Gly 2446	CCC Val	GTC Val	2403 660 Gly 2457 667	CTG Leu GAA	CCT	2412 GGT Gly 2466 GGC	CAG GLn AAA	AGA Arg CAA	2421 GGA Gly 2475 GGT	GAG Glu	ACA Arg	2430 GGC Gly Gly 2484 GGA
ATT GCT GGA CAG  Ile Ala Gly Gin  2439  TTC CCT GGT CTT.  Phe Pro Gly Leu	2394 G CGT GGT Arg Gly 2446 CCT GGC Pro Gly	GTG Val CCC	GTC Vai Vai Ser	2403 GGC GLy 2457 GGT GGT	CTG Leu GAA	CCT Pro	2412 GGT Gly 2466 GGC GGC Gly	CAG Gln AAA Lys	AGA Arg CAA Gln	2421 GGA Gly 2475 GGT Gly	Glu CCC Pro	ACA Arg TCT Ser	2430 GGC Gly 2484 GGA GLy
ATT GCT GGA CAG  Ile Ala Gly Gin  2439  TTC CCT GGT CTT  Phe Pro Gly Leu  2493	2394 G CGT GGT Arg Gly 2448 CCT GGC Pro Gly 2502	GIG Val	GTC Vai Vai Ser	2403 66C Gly 2457 66T 61y 61y	CIG Leu GA Glu	CCT Pro	2412 GGT Gly 2466 GGC Gly 2520	CAG Gln AAA Lys	AGA Arg CAA Gln	2421 GGA Gly 2475 GGT Gly 2529	GAG Glu CCC Pro	ACA Arg TCT Ser	2430 GGC Gly 2484 GGA Gly 2538
ATT GCT GGA CAG  Ile Ala Gly Gin  2439  TTC CCT GGT CTT.  Phe Pro Gly Leu	2394 G CGT GGT Arg Gly 2448 CCT GGC Pro Gly 2502	GIG Val	GTC Vai Vai Ser	2403 66C Gly 2457 66T 61y 61y	CIG Leu GA Glu	CCT Pro	2412 GGT Gly 2466 GGC Gly 2520	CAG Gln AAA Lys	AGA Arg CAA Gln	2421 GGA Gly 2475 GGT Gly 2529	GAG Glu CCC Pro	ACA Arg TCT Ser	2430 GGC Gly 2484 GGA Gly 2538
ATT GCT GGA CAG  Ile Ala Gly Gin  2439  TTC CCT GGT CTT  Phe Pro Gly Leu  2493	2394 G CGT GGT Arg Gly 2448 CCT GGC Pro Gly 2502 CGT GGT	CCC Pro	GTC Val  TCT Ser CCC	2403 660 Gly 2457 661 Gly 511 667	CTG Leu GAA Glu	CCT Pro	2412 GGT Gly 2466 GGC Gly 2520 GGC	CAG Gln AAA Lys CCC	AGA Arg CAA Gln CCT	2421 GGA Gly 2475 GGT Gly 2529 GGA	GAG Glu CCC Pro	ACA Arg TCT Ser	2430 GGC Gly 2484 GGA GLy 2538 GGA
The Ala Gly Glant CCT GGT CTT GGT CTT GGT CTT GGT CTT GGT GG	2394 G CGT GGT Arg Gly 2448 CCT GGC Pro Gly 2502 CGT GGT Arg Gly	GTG Val CCC Pro	GTC Vai Vai Ser CCC Pro	GLy GLy GLy GLy GLy GLy GLy	CTG Leu GAA Glu	CCT Pro ATG	2412 GGT Gly 2466 GCC Gly 2520 GGC GGC Gly	CAG Gln AAA Lys CCC	AGA Arg CAA Gln CCT	2421 GGA Gly 2475 GGT Gly 2529 GGA GGI Gly	GAG Glu CCC Pro	ACA Arg TCT Ser GCT	2430 GGC G1y 2484 GGA G1y 2538 GGA G1y
ATT GCT GGA CAG  Ile Ala Gly Gir  2439  TTC CCT GGT CTT.  Phe Pro Gly Leu  2493  GCA AGT GGT GAA  Ala Ser Gly Glu  2547	2394 G CGT GGT Arg Gly 2448 CCT GGC Pro Gly 2502 CGT GGT Arg Gly 2556	GTG Val CCC Pro	GTC Val TCT Ser CCC Pro	2403 GGC GLy 2457 GGI GLy 2511 GGI GGI 565	CTG Leu  GAA Glu  CCC Pro	CCT Pro ATG	2412 GGT Gly 2466 GGC Gly 2520 GGC GGC Gly 2574	CAG Gln AAA Lys CCC Pro	AGA Arg CAA Gln CCT Pro	2421 GGA Gly 2475 GGT Gly 2529 GGA GGA Gly 2583	GAG Glu CCC Pro	ACA Arg TCT Ser GCT Ala	2430 GGC Gly 2484 GGA Gly 2538 GGA GLY 2592
The Ala Gly Glant CCT GGT CTT GGT CTT GGT CTT GGT CTT GGT GG	2394 G CGT GGT Arg Gly 2448 CCT GGC Pro Gly 2502 CGT GGT Arg Gly 2556	GTG Val CCC Pro	GTC Val TCT Ser CCC Pro	2403 GGC GLy 2457 GGI GLy 2511 GGI GGI 565	CTG Leu  GAA Glu  CCC Pro	CCT Pro ATG	2412 GGT Gly 2466 GGC Gly 2520 GGC GGC Gly 2574	CAG Gln AAA Lys CCC Pro	AGA Arg CAA Gln CCT Pro	2421 GGA Gly 2475 GGT Gly 2529 GGA GGA Gly 2583	GAG Glu CCC Pro	ACA Arg TCT Ser GCT Ala	2430 GGC Gly 2484 GGA Gly 2538 GGA GLY 2592
ATT GCT GGA CAG  Ile Ala Gly Gir  2439  TTC CCT GGT CTT.  Phe Pro Gly Leu  2493  GCA AGT GGT GAA  Ala Ser Gly Glu  2547	2394 G CGT GGT Arg Gly 2448 CCT GGC Pro Gly 2502 CGT GGT Arg Gly 2556 TCT GGA	CCC Pro	GTC Val TCT Ser CCC Pro	2403 GGC GLy 2457 GCI GLY 2511 GGT GLY 2565 GGG	CTG Leu  GAA Glu  CCC Pro	CCT Pro	2412 GGT  Gly 2466 GGC  Gly 2520 GGC  Gly 2574 GCT	CAG Gln AAA Lys CCC Pro	AGA Arg CAA Gln CCT Pro	2421 GGA Gly 2475 GGI Gly 2529 GGA GGI 2583 GGT	GAG Glu CCC Pro TTG Leu	AGA Arg TCT Ser GCT Ala	2430 660 Gly 2484 664 Gly 2538 664 Gly 2592 664
ATT GCT GGA CAG  11a Ala Gly Gla  2439  TTC CCT GGT CTT.  Phe Pro Gly Leu  2493  GCA AGT GGT GAA  Ala Ser Gly Glu  2547  CCC CCT GGT GAA  Pro Pro Gly Glu	2394 CGT GGT Arg Gly 2448 CCT GGC Pro Gly 2502 CGT GGT Arg Gly 2556 TCT GGA Ser Gly	CCC Pro	GTC Val TCT Ser CCC Pro Glu	2403 660 61y 2457 61y 61y 661y 565 61y	CTG Leu  GAA Glu  CCC Pro	CCT Pro  ATG Met  CCT Pro	2412 GGT Gly 2466 GCC Gly 2520 GGC GGC GGC Ala	CAG Gln AAA Lys CCC Pro	AGA Arg CaA Gln CCT Pro	2421 GGA Gly 2475 GGI GIY 2529 GGA GIY 2583 GGI GIY	CCC Glu CCC Fro TTG Leu	AGA Arg TCT Ser GCT Ala	2430 GCC Gly 2484 GCA Gly 2538 GGA GLY CGA GLY CGA GLY
ATT GCT GGA CAG  Ile Ala Gly Gla  2439  TTC CCT GGT CTT  Phe Pro Gly Leu  2493  GCA AGT GGT GAA  Ala Ser Gly Glu  2547  CCC CCT GGT GAA  Pro Pro Gly Glu  2601	2394 CCT GGT Arg Gly 2446 CCT GGC Pro Gly 2502 CGT GGT Arg Gly 2556 TCT GGA Ser Gly 2610	GTG Pro CCC Pro CGT Arg	GCC Val  TCT Ser  CCC Pro Glu  Z	2403 650 Gly 2457 61y 5511 61y 565 61y 61y 61y	CTG Leu GAA Glu CCC Pro	CCT Pro  Arg Met CCT Pro	2412 GGT Gly 2466 GCC Gly 2520 GGC GGC GGC GGC GGC GGC GGC GGC GGC GG	CAG Gln AAA Lys CCC Pro	AGA Arg CAA Gln CCT Pro	2421 GGA Gly 2475 GGI Gly 2529 GGA Gly 2583 GGI GST GST GST GST GST GST GST GST	CCC Glu CCC Fro TTG Leu	AGA Arg TCT Ser GCT Ala	2430 6GC Gly 2484 6GA Gly 2538 6GLy 6GLy 2592 6GLy 2646
ATT GCT GGA CAG  11a Ala Gly Gla  2439  TTC CCT GGT CTT.  Phe Pro Gly Leu  2493  GCA AGT GGT GAA  Ala Ser Gly Glu  2547  CCC CCT GGT GAA  Pro Pro Gly Glu	2394 CCT GGT Arg Gly 2446 CCT GGC Pro Gly 2502 CGT GGT Arg Gly 2556 TCT GGA Ser Gly 2610	GTG Pro CCC Pro CGT Arg	GCC Val  TCT Ser  CCC Pro Glu  Z	2403 650 Gly 2457 61y 5511 61y 565 61y 61y 61y	CTG Leu GAA Glu CCC Pro	CCT Pro  Arg Met CCT Pro	2412 GGT Gly 2466 GCC Gly 2520 GGC GGC GGC GGC GGC GGC GGC GGC GGC GG	CAG Gln AAA Lys CCC Pro	AGA Arg CAA Gln CCT Pro	2421 GGA Gly 2475 GGI Gly 2529 GGA Gly 2583 GGI GST GST GST GST GST GST GST GST	CCC Glu CCC Fro TTG Leu	AGA Arg TCT Ser GCT Ala	2430 6GC Gly 2484 6GA Gly 2538 6GLy 6GLy 2592 6GLy 2646
ATT GCT GGA CAG  Ile Ala Gly Gla  2439  TTC CCT GGT CTT  Phe Pro Gly Leu  2493  GCA AGT GGT GAA  Ala Ser Gly Glu  2547  CCC CCT GGT GAA  Pro Pro Gly Glu  2601	2394 CGT GGT Arg Gly 2446 CCT GGC Pro Gly 2502 CGT GGT Arg Gly 2556 TCT GGA Ser Gly 2610 CCT GGC	GTG Val CCC Pro CCC Arg	GTC Val  TCT Ser CCC GAG GAG AAG	2403 6650 61y 2457 61y 5511 60y 565 61y 619 619	CTG Leu GAA Glu CCC Pro GCT Ala	CCT Pro ATG CCT Pro CCT Pro	2412 GGT Gly 2466 GGC GGL 2520 GGL GGL GGL GGL GGL GGL GGL GGL GGL GG	CAG Gln AAA Lys CCC Pro	AGA Arg CAA Gln CCT Pro GAA GLu	2421 GGA  Gly 2475 GGI Gly 2529 GGI GSG GSG GSG GSG GSG GSG GSG	CAG Glu CCC Pro TTG Leu TCC Ser	AGA Arg TCT Ser GCT Ala 2 CCT Pro	2430 GCC GLY 2484 GCA GLY 2538 GCA GLY 2592 GCA GLY 2592 GCA GLY 2592 GCA GCA GCA GCA GCA GCA GCA GCA
ATT GCT GGA CAG  Ile Ala Gly Gla  2439  TTC CCT GGT CTT.  Phe Pro Gly Leu  2493  GCA AGT GGT GAA  Ala Ser Gly Glu  2547  CCC CCT GGT GAA  Pro Pro Gly Glu  2601  CGA GAC GGT TCT	2394 CGT GGT Arg Gly 2446 CCT GGC Pro Gly 2502 CGT GGT Arg Gly 2556 TCT GGA Ser Gly 2610 CCT GGC	GTG Val CCC Pro CCC Arg	GTC Val  TCT Ser  CCC GAG GAG AAG	2403 65C 61y 2457 61y 5511 61y 565 61y 619 619	CTG Leu GAA Glu CCC Pro GCT Ala	CCT Pro ATG CCT Pro CCT Pro	2412 GGT Gly 2466 GGC GGL 2520 GGL GGL GGL GGL GGL GGL GGL GGL GGL GG	CAG Gln AAA Lys CCC Pro	AGA Arg CAA Gln CCT Pro GAA GLu	2421 GGA  Gly 2475 GGI Gly 2529 GGI GSG GSG GSG GSG GSG GSG GSG	CAG Glu CCC Pro TTG Leu TCC Ser	AGA Arg TCT Ser GCT Ala 2 CCT Pro	2430 GCC GLY 2484 GCA GLY 2538 GCA GLY 2592 GCA GLY 2592 GCA GLY 2592 GCA GCA GCA GCA GCA GCA GCA GCA
ATT GCT GGA CAG  Ile Ala Gly Gla  2439  TTC CCT GGT CTT.  Phe Pro Gly Leu  2493  GCA AGT GGT GAA  Ala Ser Gly Glu  2547  CCC CCT GGT GAA  Pro Pro Gly Glu  2601  CGA GAC GGT TCT  Arg Asp Gly Ser  2655	2394 CGT GGT Arg Gly 2448 CCT GGC Pro Gly 2502 CGT GGT Arg Gly 2556 TCT GGA Ser Gly 2610 CCT GGC	GTG Val CCC Pro CCC Arg	GTC Vai Vai TCT Ser CCC GAG Glu Z AAG Lys	2403 GGC GLY 2457 GLY 2511 GGY 565 GLY 619 6617 619 6617 673	CTG Leu GAA GIu CCC Pro GCT Ala GAC Asp	CCT Pro  ATG ATG Pro  CCT ATG ATG ATG	2412 GGT Gly 2466 GGC Gly 2570 GGT Ala 2628 GGT Gly 2682	CAG Gln AAA Lys CCC Pro CCC Ala	AGA Arg CAA Gln CCT Pro Glu ACC Thr	2421 GGA 	CAG Glu CCC Pro TTG Leu TCC Ser	AGA Arg TCT Ser GCT Ala 2 CCT Pro GCT Ala	2430 6GC GLy 2484 6GA GLy 2538 6GLy 6GLy 6GLy 6GLy 6GLy 1700
ATT GCT GGA CAG  Ile Ala Gly Gla  2439  TTC CCT GGT CTT.  Phe Pro Gly Leu  2493  GCA AGT GGT GAA  Ala Ser Gly Glu  2547  CCC CCT GGT GAA  Pro Pro Gly Glu  2601  CGA GAC GGT TCT  Arg Asp Gly Ser	2394 CGT GGT Arg Gly 2448 CCT GGC Pro Gly 2502 CGT GGT Arg Gly 2556 TCT GGA Ser Gly 2610 CCT GGC	GTG Val CCC Pro CCC Arg	GTC Vai Vai TCT Ser CCC GAG Glu Z AAG Lys	2403 GGC GLY 2457 GLY 2511 GGY 565 GLY 619 6617 619 6617 673	CTG Leu GAA GIu CCC Pro GCT Ala GAC Asp	CCT Pro  ATG ATG Pro  CCT ATG ATG ATG	2412 GGT Gly 2466 GGC Gly 2570 GGT Ala 2628 GGT Gly 2682	CAG Gln AAA Lys CCC Pro CCC Ala	AGA Arg CAA Gln CCT Pro Glu ACC Thr	2421 GGA 	CAG Glu CCC Pro TTG Leu TCC Ser	AGA Arg TCT Ser GCT Ala 2 CCT Pro GCT Ala	2430 6GC GLy 2484 6GA GLy 2538 6GLy 6GLy 6GLy 6GLy 6GLy 1700
ATT GCT GGA CAG  Ile Ala Gly Gla  2439  TTC CCT GGT CTT.  Phe Pro Gly Leu  2493  GCA AGT GGT GAA  Ala Ser Gly Glu  2547  CCC CCT GGT GAA  Pro Pro Gly Glu  2601  CGA GAC GGT TCT  Arg Asp Gly Ser  2655	2394 CGT GGT Arg Gly 2446 CCT GGC Pro Gly 2502 CGT GGT Arg Gly 2556 TCT GGA Ser Gly 2610 CCT GGC Pro Gly	GTG Val CCC Pro CCC Pro GCC Ala GCCT	GTC Val  Z TCT Ser  CCC GAG Glu Lys  CCT CCT CCT CCT CCT CCT CCT CCT CCT C	2403 65C Gly 2457 65L 65L 65L 65L 65L 65L 65L 65L 65L 65L	CTC Leu GAA Glu CCC Pro GCT Ala GAC Asp	CCT Pro	2412 GGTy 2466 GGC GGTy 2520 GGC GGT Ala 2628 GGT GGT GGT GGT GGT GGT GGT GG	CAG Gln AAA Lys CCC Pro GCC Ala GAG Glu	AGAA GIN CCT Pro GAA GIV ACC Thr	24211 GGA GLy 2475 GLy 2529 GGA GLy 2529 2583 GGT GLy 2583 GGT GLY 2637 GGC GGY 2637 GGC GGC GGC GGC GGC GGC GGC GG	CAG Glu CCC Pro TTG Leu TCC Ser CCC	AGA Arg TCT Ser GCT Ala CCT Pro Ala GCT Ala	2430 6Gly 2484 6GA 6Gly 2538 6Gly 2538 6Gly 2646 6Gly 2646 6GA 700 6Gly

FIG. 27D

. 2709	2718	2727	2736	2745	2754
AAG AGT GGT	GAT CGT GGT	GAG ACT GGT	८८३ ६८३ ६६३	೦೦೦ ೦೦೦ ೦೦೯	CCC GTC GGC
Lys Ser Gly	Veb Yid Ciñ	Glu The Gly	Pro Ala Gly	Pro Ala Gly	Pro Val Gly
2763	2772	2731	2790	2799	2808
CCC SCT GGC	GCC CGT GGC	CCC GCC GGA	CCC CAA GGC	೦೦೦ ೦೦೦ ೦೦೦	GAC AAG GGT
Pro Ala Gly	Ala Arg Gly	Pro Ala Gly	Pro Gla Gly	Pro Arg Gly	Asp Lys Gly
2817	2825	2835	2844	2853	2862
GAG ACA GGC	CAA CAG GGC	CAC AGA GGC	ATA AAG GGT	CAC CGT GGC	TIC ICI GGC
Glu Thr Gly	Glu Gla Gly	yzb yzd <i>G</i> yà	Ile Lys Gly	His Arg Gly	Phe Ser Gly
2871		2889		2907	
CTC CAS GGT	ಯ ಯಾ ಯ	CCI CCI GGC	TCT CCT GGT	GAA CAA GGT	CCC LCL CCF
Leu Gin Gly	Sto Sto Gy	Sto Sto Giå	Ser Pro Gly	Glu Gln Gly	Pro Ser Gly
2925	2934	· 2943	2952	2961	2970
GCC TCT GG7	CCT GCT GGT	ಯ ಯ ಹಾ	ಯ ಯ ಹ	TOT GOT GGT	GCT CCT GGC
Ala Ser Gly	Pro Ala Gly	Sto yta elà	Pro Pro Gly	Ser Ala Gly	Ala Pro Gly
2979	2988	2997	3006	3015	3024
AAA GAT GGA	CTC AAC GGT	೧೯೦ ೧೯೯೮	CCC ATT GGG	ಯ ಯಾ ಅಕಾ	ಯ ಯ ಯಾ
The Yeb Oth	Leu Asa Gly	Leu Pro Gly	Pro lle Gly	Pro Pro Gly	Sao Yad GJA
3033	3042	3051	3060	3069	3078
CGC ACT GGT					CCT CCT GGT
Arg The Gly	Asp Ala Gly	Pro Val Gly	Pro Pro Gly	Pro Pro Gly	Sio Sic CJA
3087	3096	3105	3114	3123	3132
CCC CCT GGT					_
Sto Sto GJÀ	Pro Fro Ser	Ala Gly Phe	Asp Phe Ser	Phe Leu Pro	Gln Pro Pro
3141	3150	3159	3168		
CAA GAG AAG				GCT 3'	
Gim Glu Lys .	Ala His Asp	Gly Gly Arg	Tyr Tyr Arg i	Ala	•

FIG. 27E





	HCol	ColECol	
Proline CCU	139	11	
CCC	93	12	
CCA	6	27	
CCG	0	189	
Glycine			
GGU	174	147	
GGC	97	179	
GGA	64	8	
GGG	11	12	

FIG. 30

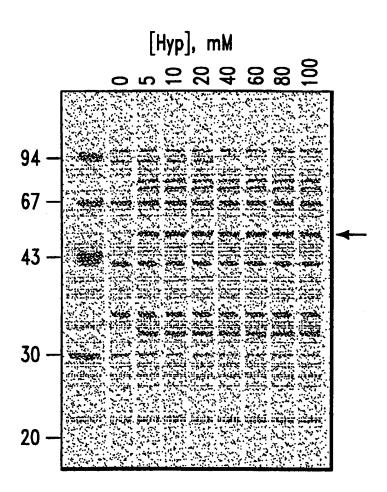


FIG. 31

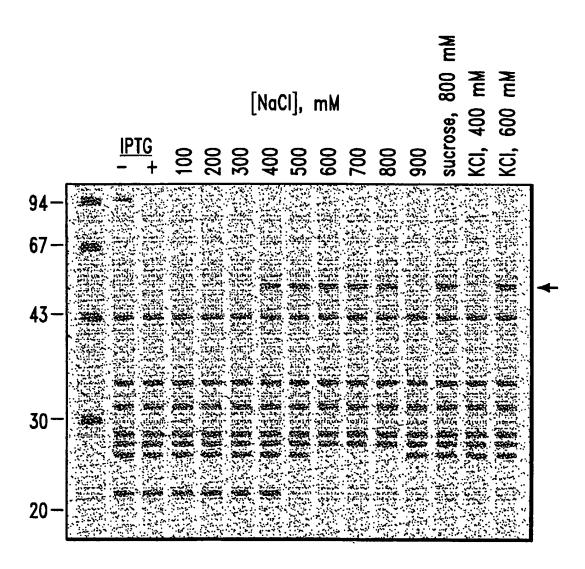
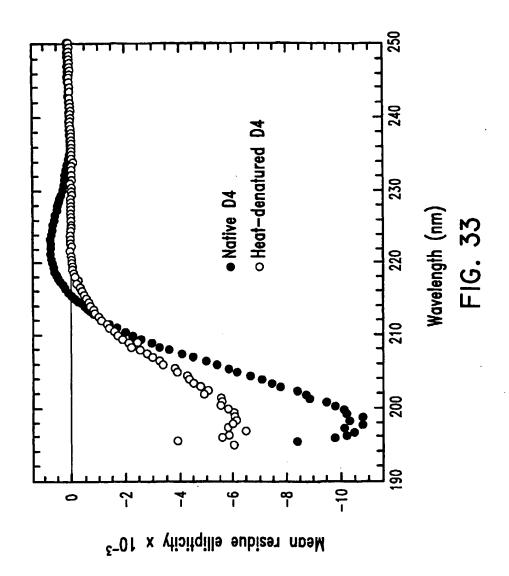


FIG. 32



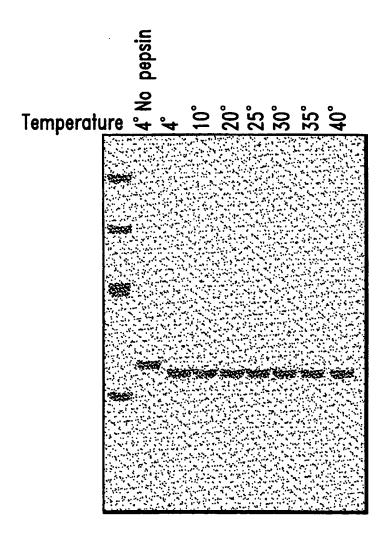


FIG. 34

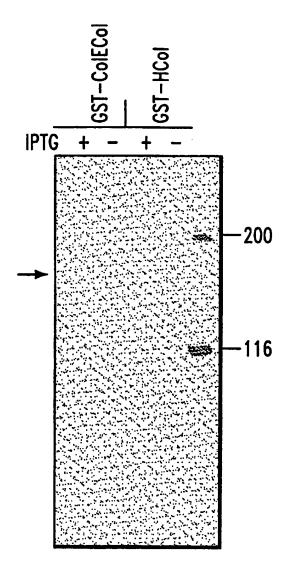


FIG. 35

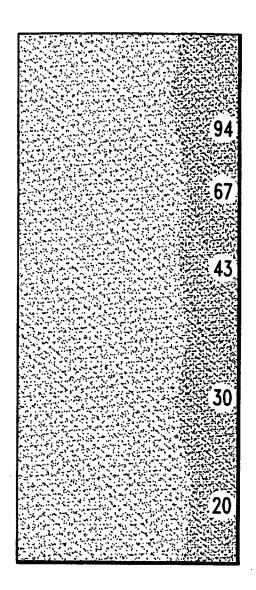


FIG. 36

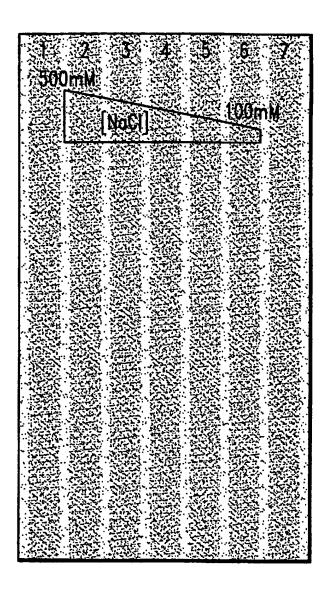


FIG. 37

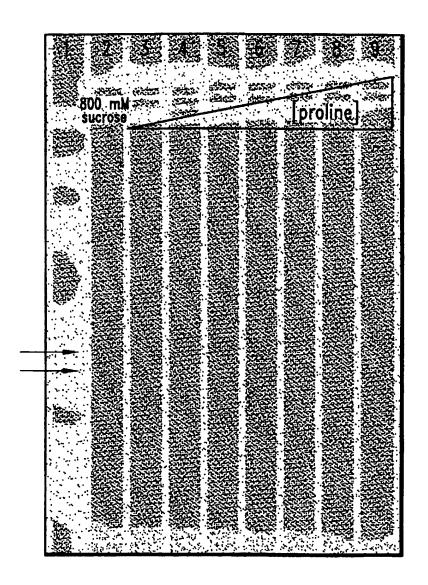


FIG. 38

5'	C)C	CTG	. 9	ТАТ	GGC	18 TAT	GAT	GλA	27 AAA	AGC	ACC	36 GGC	GGC	AIC	45 AGC		ccc	54 GGC
,																		
	Gln	Leu	Se:	Tyr	Gly	Tyr	ÇZA	Glu	Lys	Ser	Thr	CI	CIÀ	116	Ser	vai	Pro	GŢĀ
		_	63			72		^~	81	~~~		90	ccc	~``	99	ccc	ccc	108
	CCC	ATG	GGT	ccs	AGC	GGC		CGT					2					GGT
	Pro	Mec	Gly	Pro	Ser	Gly	Pro	Arg	GJA	Leu	Pro	Gly	Pro	Pro	Gly	Ala	Pro	Gly
			117			126			135			144			153			162
	CCG	CAG	GGC	TTT	CAG	GGT	$\infty$	ccc		GAA	ccc	GGC	GAA	CT	CGT	CCC	AGC	GGC
	Pro	Gln	Gly	ehe	Gln	Gly	Pro	Pro	Gly	Glu	Pro	Gly	Glu	Pro	Gly	Ala	Ser	CJÀ
			171			180			189			193			207			216
	CCG	ATG	GGC	ಯ	CGC		CCG	ccc			CCA		AAA	AAC		GAT	GAT	GGC
	Pro	Mec	GTĀ	510	Arg	GīŊ	510	PIO			Pro		rys	AS.1		νςΰ	мsр	GΣΥ
	~~~		225	224	m:	234	CT	CCC	243		CGT	252	œ	αG	261	ന്നു	CNG	270 GGC
	GAA			A														
	Glu	Ala	Gly	Lys	Pro	Gly	Arg	Pro	CJ À	Glu	Arg	Gly	Pro	Pro	GŢĀ	Pro	Gln	Gly
			279	,		288			297			305			315			324
	ထေ	CGC	GCA	CIG	<u></u>	GGT	ACT	CCG	GGA	CTG	<u> </u>	665	ATG	AAA	GGC	CAC	œc	GGT
	Ala	Arg	CJÀ	Leu	Pro	GŢĀ	Thr	Ala	Gly	Leu	Pro	GŢĀ	Met	Lvs	Gly	His	Arg	Gly
			333			342			351			360			369			378
	TTC	TCT	GT	CTG	CAT	GGT	GCC	AAA	CCA	GAC	CCC	GGT	CCG	ಯ	GGT	ငငင	aaa	GGT
	Phe	Ser	Gly	Leu	Asp	Gly	Ala	Lys	Gly	Asp	λla	Gly	Pro	Ala	Gly	Pro	Lys	Gly
			387			396			405			414			423			432
	GAG	CCG		ACC	CCG		G۶٩	AAC		CCC	ccc		CAG	atg		CCG	ŒŢ	
	Glu	Pro	Glv	Ser	Pro	Glv	Glu	Asn	Glv	Ala	220	Glv	Gln	ᄣ	Glv	Pro	Arg	Glv
						_	,					_			477		-	_
	CTG	CCT	441 GGT	GAA	œ	450 GGT	CGC	œ	459 CCC	GΩ	œ	463 GGC	CCA	GCT.	-	GCA	ŒĨ	486 GGC
	Leu	Pro	СТÅ	Glu	Arg	GIY	Arg	Pro	GIY	ALA	Pro	GīĀ	Pro	ALA	GTĀ	Ala	Arg	GLY
		C>#	495	~~	×~	504	~~	~~	513	œ,	~~	522	~~	300	531	~~~	~~	540
	AAC	GAT			MUL				901		·					CCG		GGT
	Asn	Asp	Gly	Ala	Thr	Gly	Ala	Ala	CJÀ	Pro	Pro	GJÀ	Pro	Thr	GŢĀ	Pro	Ala	Gly
			549			558			567			576			585			594
	ccc	CCC	Œ	TTT	∞	GGT	ငေ	CLC	GGT	ထေ	AAA	GGC	GAA	GCA	GGT	ccc	CAG	GGG
	Pro	Pro	Gly	Phe	Pro	Gly	Ala	Val		Ala	Lys	CJĀ	Glu	Ala	Gly	Pro	Gln	Gly
			603			612			621			630			639			648
	CCG	CGC		AGC	CAG		CCT	CYC		GTT	CCT		GAA	CCG		CCG	∞	
	Pro	Arc	Gly	Sar	Glu	Glv	Pro	Gln	Glv	Va)	Arc	Glv	Glu	Pro	Glv	Pro	Pro	G) v
		y		J-12	510			J	Ī		9		J_u	•				-
	ccc	GCG	657 GGT	ca:	GCG	666 GSC	222	GCT	GGT	AAC	CCT	GGC	GCG	CAC	693 GGT	CAG	CC A	702 GGT
		·																
	Pro	Ala	Gl'n	Ala	Ala	Gly	Pro	Ala	GΙΆ	Asn	Pro	GŢĀ	Ala	eş.	Gly	Gln	Pro	Gly

FIG. 39A

					7.17	756
711 CKIG AAA GGT G	720	ons one et	29 TO ATT GC	827 A GGT GCA	ಯ ಯ	
CACC AVV CCL C						She Pro Gly
Ala Lys Gly A	Ts yeu Già	Ala Pro G	ly He Al	a Giv Ala	SEO GIÀ	
765	774	7(83	792	801	610 666 666
765 GCC CGC GGC C	ಯ ಉದ್ದರ್	CCC CYC C	cc ccc cc			
Ala Arg Gly F	Pro Ser Giy	Pro Gln G	ly Pro Gl	y Gly Pro	Fro Gly	Fro Lys Gly
		٥	27	845	855	864
ALC AGC GGT (CAA CCG GGT	GCC CCC G	GC AGC AA	A GGC GAC	YCC COL	GCG AAA GGT
Asn Ser Gly	Glo Pro Gly	Ala Pro G	iy Ser Ly	s Gly Asp	THE GLY	Ala Lys Gly
	222	8	191	900	909	918
Gar coe eec (CCL GTG GGT	GTT CAA G	သ သာ ဘိ	e eec cce	CCG GCC	CAG GAA GGC
Glu Pro Gly						
	_		45	954	963	972
927 AAA CGC GGT (ಣೀ ಪಾ ಯ ಪಾ	GAA CCC G	200 COC YC			
Lys Arg Gly						
•					1017	
981 : GAN CGT GGT !	. 990 GGC CGG GGT	AGC CCC G	999 337 777 CC	1003 16 GGC GCG		
Glu Arg Giy						
GIR WEG GIA						
1035 CCG AAA GGT (1944 152 SCS SCI		153 STACCO	1062 G GGC CCG	1071 CCG GGC	1080 CCA AAA GGC
	'		- -			
Sto The Cy	hto vis età	GIC Arg G	iy ser er	o GTÅ 520	Aid GIY	SEO TAR CIA
1089 AGC CCG GGC (8991 גרפ ברט פנס		07 GT 632 GO		1125	1134 CCC 333 CCT
Ser Pro Gly (Gir Ala Giy	Arg Pro G	ly Glu Al	a Gly Leu	Fro Gly	Ala Lys Gly
1143	11.52	11		1170	1179	1188
CTG ACC GGC	11. UU 44.	AGC 003 G	01 CT CA	r GGC AAA		505 CC. GGT
Leu Thr Gly S	Ser Pro Gly	Ser Pro G	ly Pro As	o Gly Lys	Thr Gly	Szó Szo CJA ·
i197	1206	12		1224	1233	1242
000 000 000 0	CAG GAT GGT	000 000 0	sc ccs cc	3 665 666	CCG GGT	000 007 00T
Pro Ala Gly (Gin Asp Gly	Arg Pro G	ly Pro Pro	Gly Pro	Pro Gly	Ala Arg Gly
. 1251	1250	. 126	69	1278	1287	1296
CAG GCG GGT G	one and sec	TTT CCA CC	3C COC AN	A GGT GGG	ടോ ടോ	GAA CCC CGC
Gin Ala Gly V	Val Met Gly	Phe Pro G	ly Pro Ly	Gly Ala	Fla Gly	Glu Pro Gly
1305	1314	132	23	1332	1341	1350
AAA GCG GGC C				GCC GCT		
Lys Ala Gly (Glu Arg Gly	Val Pro G	ly Pro Pro	Gly Ala	Val Gly	Pro Ala Gly
1359	1368	137		1356	·1395	1404
AAA GAT GGC G						
Lys Asp Gly G	Slu Ala Glv	Ala Gln Gl	ly Pro Pro	Gly Pro	Ala Glv	Pro Ala Glv
- •	-		_	•		

FIG. 39B

						.					,	440		,	458
1413 GAG COC GGT		1	422		CC2.	431	2Ω°	ccs.	GGT	770	CAG	GGT	CTG		
GAG CGC GGT	GAA	<u>_4</u>													 Cl.,
Glu Arg Gly	Glu	Gln	GΪΆ	9ro	Ala	ВЗ	Ser	623	Gly	?:.e	Gin	GIA	Leu	210	GrA
									404		,	503		3	512
1467 CCT GCG GGT	CCA	ငင္ေ	GGT	GAA	GCS	GCC	λλλ	CCG	ဏ	GΥY	CAA	GGT	GTG	œ	GGC
Pro Ala Gly				Glu	212	Glv	Lvs	Pro	Glv	GLu	Gln	Gly	Val	250	Gly
Pro Ala Gly	Pro	Pro	GIA	GIG											1566
1521 GAC CTG GGC			1530	~~	300	1539	ccc	CCC	1548 GGC	GAA	œσ	1557 GGT	TTC		
GAC CTG GGC	GCC	CCA													
Asp Leu Gly	, Ala	Pro	Gly	Pro	Ser	GŢĀ	Ala	Arç	Gly	Gžu	Arg	GIÀ	5.re	210	GTĀ
						1 502			1 602			1611			1620
1575 GNA CGT GG	GTC	CAG	GGC	CCG	œ0	Œ	∞	CC:	GGT	COS	CGC	GGC	GCC	YVC	GGC
Glu Arg Gly															
GIU Arg GI	, va.														1674
1629 GCG CCG GG	3		1638 661	CO	: 222	1647	GAT	CCG	1656 GGT	GCC		1665 .GGT			
Ala Pro Gl	y Asr	gzA ı	Gly	Ala	Lys	Gly	Y20	Ala	Gly	a <u>'</u> a	Pro	Gly	Ala	Pro	GIĀ
168	3 •		1692	:		1701			1710			1719			1725
AGC CAG GG	cα	CCC	GGC	CTG	CAR	GGC	ATG	CCG	GGT	CAA	CGT	GGT	CCC	GCG	GGT
Ser Gln Gl	· Ala	. Pro	Glv	Leu	Gla	Gly	χet	Pro	Gly	, CLu	Arg	Gly	Ala	Ala	Gly
173° CTA CCG GG	ו הרכנ		1746 GGC			1755 GGT			1764 GGT			1773 . GGT			1782 GGC
	- '				·										
Leu Pro Gl	/ Pro	Lys	Gly	ysa	Arg	GJĀ	, yzb	<u> </u>	Gly	?:0	Lys	GŢĀ	Ala	Asp	Gly
179			1800						1918			1827			1836
TCC CCT GG	AA	CAT	GGC	GII		CCT	CTG	λCC	GGC	ಯ	ATC	GGC	<u></u>	ccc	GGC
Ser Pro Gly	/ Lys	Asp	Gly	Val	Arg	Gly	Leu	The	Gly	Pro	Ile	Gly	Pro	Pro	Gly
184			1854			1863			1872		,	1881		,	1890
CCC GCA GG				•						ಯ			CCA		
						~~~	~~~		<b></b>						
Pro Ala Gly	ALA	Pro	GTĀ	ASO	LÿS	шħ	GIU	261	CTĀ	# <b>T</b> 0	Ser	CTA	PTO	ALA	CTÅ
1899			1908			1917		1			-	1935		-	944
CCC ACT GGT	GOG	Œī	GGT	GCC	œ	GGC ———	GAC	<u> </u>	GGT	CHA	<b>CCG</b>	GGT	CC3	ccc	GGC
Pro Thr Gly	, Ala	Arg	Gly	Ala	?ro	Gly	Asp	Arş	Gly	GLu	Pro	Gly	Pre	Pro	Сĵå
_: 1953	ı		1962			1971		1	980		1	9RQ		1	.998
CCC GCC GCC	TIT	œ	GGC	CCG	$\alpha$	GGC	GCT	CAC	GGC	C÷G	$\infty$	GGI	GCG	AAA	
Pro Ala Gly			614	÷	2		112			 Cln	D-0		11.	~'	
sto wa m	FILE	nia	GLY	810	:10	ary	Αια	wż	GrÅ	Gill	PLO	Gry	VIG	råz	GIĀ
2007			2016			2025	C)C		034	ccc		2043	<b>ب</b>		2052
GAA CCG GCC					~~A					ولنا 					
Glu Pro Gly	' Asp	Ala	Gly	Ala	Lys	СſĀ	ÇZÁ	Αlε	Gly	2:0	Pro	Gly	Pro	Ala	Gly
2061		:	2070		:	2079		2	2088		2	2097		2	106
CCC CCC CCC							AAC	G£3	GGT	CCG			CCC		
Pro Ala Gly	Pro	Pro	Glv	Pro	Ile	Glv	Asn	 !sV	Glv	21a	Pro	Glu	Ala	Lve	G) v
			3			,		_	1			1			1

FIG. 39C

									~ 1 4 3			2:51			2160
2115 GCG CGC GGC F	, CC -	ت-ب 2	124	CCG	ccs	2133 GGC	ထေး	ACC	2142 GGT	770	CCC				
Ala Arg Gly	Ses	ΥŢα	Gl'n	510	613	GLY	ة شد	TES	arā	Fe	520	ودي			
2169		2	178			2187		~~	2196		~	2205 ~~~	cca		2214
CGC GTG GGT (															
Arg Val Gly	2:0	?ro	Gly	Pro	Ses	Gly	Asn	Ala	Gly	Pro	Pro	Gly	bro	520	Gŗà
0003			222			2241			2250			2259		•	2258
CCC CCC CCC 2	eaa	ಎಲ	GGC	GCC	AZA	GGT.	$\infty$	CGI	GGT	€÷≯	ACC	933	CCI	GCG	GGA
Pro Ala Gly	Lvs	Glu	Gly	Gly	Lys	Gly	Pro	Arç	Gly	Glu	The	Gly	Pro	Ala	Gly
			- 2285			2295			2304			2313			2322
CGT CCA GGT	GRA	GTG	GGT	CCG	CCG	GGC	ccc	CCG	GGC	ಯ				AAA	GGT
Arg Pro Gly															
•	<u> </u>							•							- 2376
2331 AGC CCG GGT	ထေ	GAT	2340 GGT	$\infty$	GCC	2349 : GGT	305		2358 . GGC			2367 GGT			
				~											
Ser Pro Gly	YIS	.Sp	GIĀ	Pro	A_C	Giy	Ald	PES	GIY	1-15					
2395 ATC GCT GGC	CSG.		2394 CGT			2403			2412 . GGT			2421 333			2430 GGC
Ile Ala Gly	GŢĽ.	Arg	G17	Val	Val	Gly	Leu	Pro	Gly	Gin	Arg	Gly	Glu	Arg	GīĀ
2439			2448	~~~		2457	cic		2466			2475	~		2484
TTT CCG GGT					٠	691									
Phe Pro Gly	Leu	Pro	Gly	Pro	Ser	Gly	Glu	Pro	Gly	Lys	Gin	Gly	Pro	Ser	GŢĀ
2493			2502			2511			2520			2529			2538
GCG AGC GGT (	G÷7	CGT	GGC	CCG	œ	GGT	<u>acc</u>	AIG	GGC		<u> </u>	GGT	CTG	GCG	ccc
Ala Ser Gly	Glu	Arg	Gìy	Pro	Pro	Gly	520	Me:	Gly	210	Pro	Gly	Leu	Ala	Gly
2547		2	2556		:	2565		2	2574		2	583		2	2592
CCT CCG GGT (	CY?	AGC	GGT	CGT	CHA	GGC	GCC	CC3	ಡಾ	CCC	Gła	œ	AGC	CCA	GGC
Pro Pro Gly (	Glu	Ser	Gly	Arg	G1u	Gly	Ala	Pro	Gly	الم	Glu	Gly	Ser	Pro	Gly
2601		,	610		,	2619		-	628		2	637		-	646
CCC CAC CCT A	AGC (			$\infty$						GRA			$\infty$		
Arg Asp Gly S	Ser!	Pro	Glv	Ala	Lvs	Glv	ASD	Are	Glv	Glu	Thr	Glv	Pro	Ala	Cly
									-			_			-
2655 CCC CCG GGT (			664 GGC			2673 GGT					2 GIG		ccs		700 GGC
Pro Pro Gly A		713	Grà	MIG		•	AL A	813	GIĀ	710	ver '	orà	FIO	ALC	GIÀ
2709 AAA AGC GGT (	EAT (		71E GG7	GAG		2727 CGT	ന		735 GGC	CCC	2 (300)	745. GGT	cce	2 GTG	754 GGC
Lys Ser Gly A	rà i	Arg	Gly	Glu	Tar	Gly	?10	Aia	GLY	?:0	SIA	Gly	5ro	Val	Gly
2763	~~ <i>.</i>		7.72	ccc		781	~~		790	~~		799	~~~		808
OCA GOG GGC G	(				ال 	 G01					) وکیل 	GGT 	نبن 	A44	CGT
Pro Ala Gly A	la i	Arg	GŢ.Ā	2ro	Ala	Gly	?ro	Gla	Gly	2:0	Yrd (	Gly	Αsp	Lys	Gly

FIG. 39D

																	2862
GAA	ACG	GGC	GAA	CAG	GGC	GAL	CGI		AT	. AA	A (~)	. L.40	ابنا ۔	GG	TTC	AGC	GGC
Glu	Th	GLy	/ Glu	Gln	Gly	Asp	Arg	Gly	Ile	Ly:	s Gly	/ His	Arg	, Gly	Phe	Ser	Gly
		2871	_		2880	)		2889	)		2898	3		2907		;	2916
CTG	CYC	GG1	CCA	CCG	GGC	. ccc	CCC	GGC	AG.	cc	G GG1	CA	CAG	GGT	CCC	TCC	GGA
	C1-						D										
Leu	GIT	, CT	Pro	Pro	GTA	Pro	Pro	GIY	. 26:	720	o Gil	, GIC	ı Gin	GIA	Pro	Ser	Gly
			5		2934			2943			2952	:					2970
GCC	AGC	CGC	; ccc	GCG	GGC	CCA	CCC	GGT	ca	CCC	GGC	: AGC	CCC	GGT	GCG	CCG	GGC
				22-	C1.		3	Cl.,									
ALA	262	GIÃ	PIO	WTG	GTĀ	910	wrg	GIÀ	PES	PEC	o etă	Ser	. Ala	GIV	Ala	Pro	Gly
		2979	1		2988			2997			3006	;		3015		7	3024
AAA	೧೯೦	CĊI	CIG	AAC	GGT	CTG	CCC	GCC	œ	ATC	GGC	CCG	$\infty$	CCC	CCA	ထင	GGC
							~~~										
Lys	ريد	Gry	Leu	W2!!	GIĀ	Leu	PEC	GTĀ	YEO	TTE	GLY	Pro	?ro	Gly	Pro	Arg	GLy
					3042			3051			3060			3069		•	078
CCC	ACC	CCT	CAT	GCG	GGT	∞	GTG	CCT	∞	CCG	GC	000	∞	æ	ထာ	CCA	GGC
λty	11.2	GIĀ	κŞ	ALE	GIA	PIO	AST	Gly	Pro	Pro	Gly	Pro	Sto	Gly	Pro	Pro (Gly
	;	3087	. '	3	3096		3	3105			3114		,	1122		3	
CCC	CCG	CCT	∞	CCC	AGC	CCC	GGT	TTC	G¥€	TTC	AGC	TTC	cre	æ	CAG (.د م	132
		CLY	*10	FLO	zet	MIG	CTÀ	rne	Asp	Phe	Ser	Phe	Leu	Pro	Gln i	Pro E	ro
	3	1141		3	150		,	160		_							
CAG (G+G	FAA	CCG	CAC	CAC	GGC	GGT	CGC	TAC	TAC	CET	GCG	3 '				
													-				
Gln (-10			ا باتد	ara (ath 1	arg '	Tyr	Tyr	<u>λ=g</u> .	Ala					

FIG. 39E

EcoR1 start Oligo N1-1

AGCGTGCCCCCCCGATCGGTCCGACC-3'

3'-GGCCCGGGGCCCACCGGGCCCACCGGGCCCCCGGGGTCCAGCGGGGGCCACCATTATTCGAACCC-5'

Oligo N1-2

EcoR1 Rsr II Oligo N1-3

3'-TACCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGTTTTTGCCGCTACTACCGCTTCGCCCGTTTGCCAGGCATTATTCGAACCC-5'

Oligo N1-4

FIG. 40

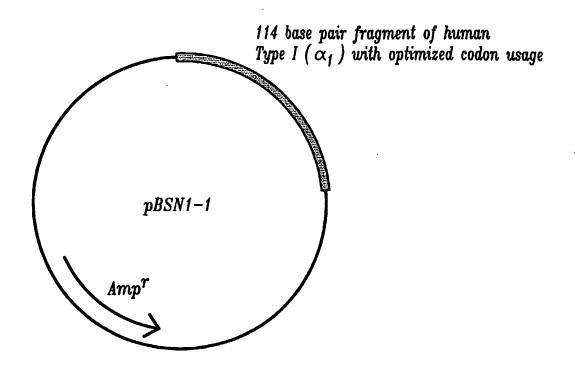


FIG. 41

5.					GGC		GAT	AFD		AGC 	ACC	GGC	<u> </u>	ATC		<u> </u>		54 6GC
	Gin	Leu	Ser	Tyr	CIA	ryr	ASP	GIU	Lys	Ser	TUE	GTA	Q+3	176	ser	٧ <u>۵</u> _	220	G÷À
			63			72			81			90		٠	59			108
	ಯ	ATG	GGT	CCG	AGC	CCC	∞	CGT	GGC	CIG	∞	GCC	α	CCA	337	ಯ	α	GJT
																	,	
	Pro	Met	Gly	Pro	Ser	Gly	Pro	Arg	Gly	Leu	Pro	Gly	620	Pro	31À	A) a	Pro	GŢĀ
	Ω್																	
	520																	

FIG. 42

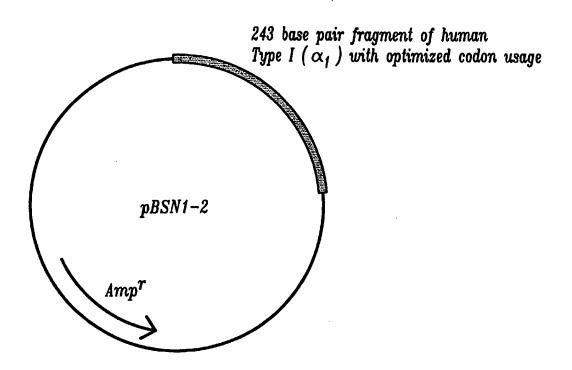


FIG. 43

					18										54			
5,	CAG	C:G	AGC	TAT	œ	TAT	GAI	GAA	AAA	AGC	ACC	GGC	GGC	ATC	AGC	CTG	CCG	GGC
	CJV	Leu	Ser	Tyr	Gly	Tyr	Asp	Glu	Lys	Ser	The	Gly	Gly	Ile	Ser	Val	9.70	Gly
			63												99			108
	∞	ATG	GGT	CCG	AGC	GGC	ccc	CGT	GGC	CTG	∞	GGC	ccc	CCA	GGT	α	α	GGT
	Pro	۲	Gly	510	Ser	Gly	Pro	Arg	Gly	Leu	Pro	Gly	Pro	Pro	Gly	λla	Pro	Gly
			117			126			135	•		144			153			162
	ငင္သာ	وج	GGC	TTT	CAG	GGT	∞	CCC	GGC	CAA	œ	GGC	GAA	CCT	CGT	CCG.	AGC	GGC
	Pro	GŢV	Gly	Phe	Gln	Gly	510	Pro	Gly	Glu	Pro	Gly	Glu	Pro	Gly	Ala	Ser	Gly
			171			180			189			198			207			216
	ccc	ATG	GGC	CCG	ССС	GCC	CCG	∞	GGT	œ	CCA.	GGC	AAA	220	CCC.	CET	CET	510
																CAL	<u> </u>	GGC
	Pro	たって	Gly	?ro	Arg	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Lys	As::	Gly	ನಿತ್ತಾ	czÁ	Gly
			225			234												
	GAA	∞ G	CCC				ŒT	∞										
	Glu	λla	СſΑ	Lys	Pro	СſΑ	Arg	Pro										

FIG. 44

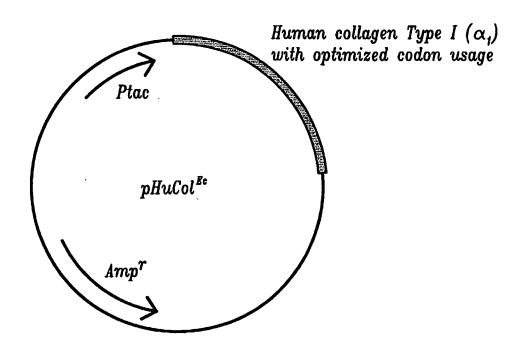


FIG. 45

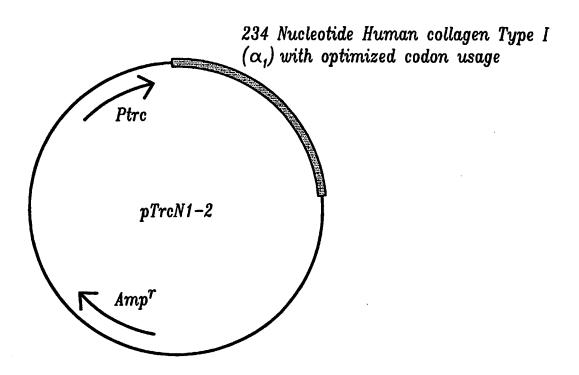


FIG. 46

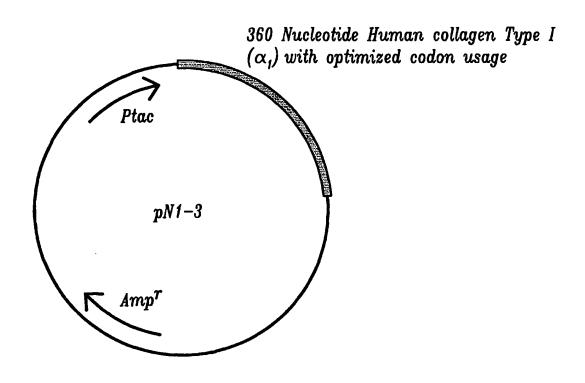


FIG. 47

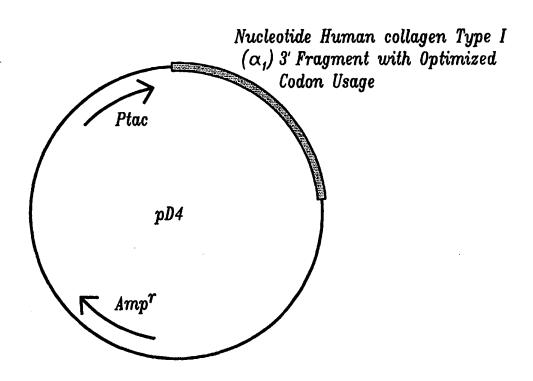


FIG. 48

9 18 - 27 36 45 54 5 CAG TAT CAT GGA MAN GGA GTT GGA CTT GGC CCT GGA CCA ATG GGC TTA ATG GGA
were near that many house the time and the time that the time that the time the
Gln Tyr Asp Gly Lys Gly Val Gly Leu Gly Pro Gly Pro Met Gly Leu Met Gly
63 72 81 90 99 108 OCT AGA GGC CCA CCT GGT GGA GCC GGA GGC CCA GGC CCT CAA GGT TTC CAA GGA
Pro Arg Gly Pro Pro Gly Ala Ala Gly Ala Pro Gly Pro Gln Gly Phe Gln Gly
117 126 135 144 153 162 OCT GCT GCT GCT GCT GCT GCT GCT GCT GCT G
Pro Ala Gly Glu Pro Gly Glu Pro Gly Gln Thr Gly Pro Ala Gly Ala Arg Gly
171 180 189 198 207 216 CCA GCT GGC CCT CCT CGC AAA CCC GGA
Pro Ala Gly Pro Pro Gly Lys Ala Gly Glu Asp Gly His Pro Gly Lys Pro Gly
225 224 242 242
225 234 243 252 261 270 CGA CCT GGT GGG AGA GGT GTT GTT GGA CCA CAG GGT GCT CGT GGT TTC CCT GGA
Arg Pro Gly Glu Arg Gly Val Val Gly Pro Gln Gly Ala Arg Gly Phe Pro Gly
279 288 207 206 215
ACT OCT GGA CTT OCT GGC TTC AAA GGC ATT AGG GGA CAC AAT GGT CTG GAT GGA
Thr Pro Gly Leu Pro Gly Phe Lys Gly Ile Arg Gly His Asn Gly Leu Asp Gly
333 342 351 200
THE AME GEA CAG OCC GET GCT CCT GET GTG AME GET GAA CCT GET GCC CCT GET
Leu Lys Gly Gln Pro Gly Ala Pro Gly Val Lys Gly Glu Pro Gly Ala Pro Gly
387 396 405
GAR ART GGA ACT COA GGT CAR ACA GGA GCC CGT GGG CTT CCT GGT GAG AGA GGA
Glu Asn Gly The Pro Gly Gln The Gly Ala Arg Gly Leu Pro Gly Glu Arg Gly
441 450 460
CAT GET GET GET GET GET GET GET GET GET GE
Arg Val Gly Ala Pro Gly Pro Ala Gly Ala Arg Gly Ser Asp Gly Ser Val Gly
495 504 510
CCC GTG GGT CCT GGT CCC ATT GGG TCT GCT GCC CCT CCA GGC TTC CCA GGT
Pro Val Gly Pro Ala Gly Pro Ile Gly Ser Ala Gly Pro Pro Gly Phe Pro Gly
549 558 567 576 585 594
CCC CCT GGC CCC AAG GGT GAA ATT GGA GCT GTT GGT AAC GCT GGT CCT GGT
Ala Pro Gly Pro Lys Gly Glu Ile Gly Ala Val Gly Asn Ala Gly Pro Ala Gly
CCC GCC GGT CCC CGT GGT GAA GTG GGT CTT CCA GGC CTC TCC GGC CCC GTT GGA
Pro Ala Gly Pro Arr Gly Cly Val Cly Val
Pro Ala Gly Pro Arg Gly Glu Val Gly Leu Pro Gly Leu Ser Gly Pro Val Gly
CCT CCT CGT AAT CCT CGA GCA AAC CGC CTT ACT CGT CCC AAG CGT CCT CCT CGC
Pro Pro Gly Asn Pro Gly Ala Asn Cly Isa To Co.
Pro Pro Gly Asn Pro Gly Ala Asn Gly Leu Tar Gly Ala Lys Gly Ala Ala Gly

FIG. 49A

											770			747			756
~~	~~	711 GGC (GTT 4	CCT	720 GGG	GCT	∞	729 GGC	CTC	œī	GGA	∞	œc	GGT	TTA	CCT	GGC
CIT						Ala	 Pro	Glv	 Ieu	Pro	Gly	Pro	Arg	GJA	11e	Pro	GŢĀ
											. 202			801			810
~~	C-T-eF	765 CCT	CCT	œ	774 GGT	GCT	ACT	783 GGT	œ	AGA	GC ₂	CTT	GTT	GGI	GAG	α	GGT
	G11				Clu		Thr	Glv	Ala	Arg	Gly	Leu	Val	Gly	Glu	Pro	GJ Å
											946			855			864
CCA	CCT	819 GGC	TCC	AAA	828 GGA	GAG	AGC	837 GGT	AAC	AAG	GGT	GAG	ccc	GGC	TCT	CCI	GGG
			Sar	Tue		Glu	Ser	Gly	Asn	Lys	Gly	Glu	Pro	Gly	Ser	Ala	Gly
					000			201			900	1		909			918
œ	CAA	873 GGT	CCT	œ	892 GGT	∞	AGT	GGT	GAA	GAZ	GCA	AAG	AGA	GGC	CT	AAT	GGG
																	Gly
FIC	, G11	927			936			945			954			963			972
GAF	CC1	GGA	TCI	œ	GGC	ccı	œ	GCA	c ccı	007	. CO	CIG	AGI	GGI	AGT	007	GGT
Glu	Ala	Gly	Ser	Ala	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Leu	Arg	, Gly	Ser	Pro	Gly
		` QR1			990)		999)		1008	}		1017	,		1026
TC	con	GGI	CII	· œ	GG2	GCT	GAT	GGC	AG:	(œ	. eac	GTC)TA	. eec	CT	<u> </u>	GGT
Se	Arg	Gly	Lev	Pro	GJ	Ala	AS:	Gly	Arg	Ala	a Gly	Val	. Met	Gly	Pro	Pro	Gly
		1035			1044			1053			1062			1071			1080
AG	r Œ	GGT	GCA	AĞI	GCC	C C T	GC	GGA	GIC	<u> </u>	A GGP	œ1	' AA1	GGA	GAT	GCI	GGT
Ser	: Arg	Gly	Ala	Sex	Gly	Pro	Ala	Gly	Val	Arc	g Gly	Pro	Asr	Gly	Asp	Ala	Gly
		1089			1098									1125			1134
<u> </u>	: cc1	GGG	GA.C	<u> </u>	GGI	. cro	ATC	GGA		AG	GGI	CTT	<u>~~1</u>	GGT	TCC	CC1	GGA
Arg	Pro	Gly	Glu	Pro	Gly	Leu	Met Met	Gly	Pro	Arg	Gly	Leu	Pro	Gly	Ser	Pro	Gly
		1143			1152		,	1161			1170		~~	1179			1188
AAI	JIA '			GC1		. AAA	GAA		- C1	GIC			<u>uci</u>		AIC	GAL	GGC
Asn	Ile	Gly	Pro	Ala	Gly	Lys	Glu	Gly	Pro	Val	. Gly	Leu	Pro	Gly	Ile	Asp	Gly
300	. ~~	1197			1206			1215		aca.	1224			1233			1242
				. .													GGA
Arg	Pro	Gly	Pro	Ile	Gly	Pro	Ala	Gly	ALa	Arg	r Gly	Glu	Pro	Gly	Asn	Ile	Gly
4-47.		1251			1260			1269		٠ 🖳	1278			1287	CAT		1296 GGT
Phe	Pro	Gly	Pro	Lys	GIÀ	Pro	Thr	GTA	ASE	Pro		-	ASN	Gly	Asp	Lys	Gly
CAT	GCT	1305 GGT		CT	1314 GG1			1323 GGT		· 002	1332 GGT			1341 GGA			1350 GGT
																	Gly
n s	, ALA	_													ASI1		
GCT	CAG	1359 GGA			1368 GGA			1377 GGT		CAA	1386 GGT			1395 GGT	GAA		GGT

FIG. 49B

Ala Gln Gly Pro Pro Gly Pro Gln Gly Val Gln Gly Gly Lys Gly Glu Gln Gly 1449 1440 CCC GAT GET CCT CCA GET TTC CAG GET CTG CCT GET CCC TCA GET CCC GET Pro Asp Gly Pro Pro Gly Phe Gln Gly Leu Pro Gly Pro Ser Gly Pro Ala Gly 1503 1494 1485 GAA GTT GGC AAA OCA GGA GAA AGG GGT CTC CAT GGT GAG TTT GGT CTC OCT GGT Glu Val Gly Lys Pro Gly Glu Arg Gly Leu His Gly Glu Phe Gly Leu Pro Gly 1548 CCT GCT GGT CCA AGA GGG GAA CGC GGT CCC CCA GGT GAG AGT GGT GCT GCC GGT Pro Ala Gly Pro Arg Gly Glu Arg Gly Pro Pro Gly Glu Ser Gly Ala Ala Gly 1593 1602 1611 OCT ACT GGT OCT ATT GGA AGC OGA GGT OCT TCT GGA COC OCA GGG OCT GAT GGA Pro Thr Gly Pro Ile Gly Ser Arg Gly Pro Ser Gly Pro Pro Gly Pro Asp Gly 1647 . 1665 1638 AME AME GET GAM CET GET GTG GTT GET GET GET GEC MET GET GET GET TEN TET GET Asn Lys Gly Glu Pro Gly Val Val Gly Ala Val Gly Thr Ala Gly Pro Ser Gly 1692 1701 CCT AGT GGA CTC CCA GGA GAG AGG GGT GCT GCT GGC ATA CCT GGA GGC AAG GGA Pro Ser Cly Leu Pro Gly Glu Arg Gly Ala Ala Gly Ile Pro Gly Gly Lys Gly 1746 1755 GAA AAG GGT GAA OCT GGT CTC AGA GGT GAA ATT GGT AAC OCT GGC AGA GAT GGT Glu Lys Gly Glu Pro Gly Leu Arg Gly Glu Ile Gly Asn Pro Gly Arg Asp Gly 1809 SCT OCT GGT GCT CAT GGT GCT GTA GGT GCC GCT GGT GCT GCA GCC ACA GGT Ala Arg Gly Ala His Gly Ala Val Gly Ala Pro Gly Pro Ala Gly Ala Thr Gly 1854 1863 Asp Arg Gly Glu Ala Gly Ala Ala Gly Pro Ala Gly Pro Ala Gly Pro Arg Gly 1926 AGC OCT GGT GAA OGT GGC GAG GTC GGT OCT GCT GGC CCC AAC GGA TIT GCT GGT Ser Pro Gly Glu Arg Gly Glu Val Gly Pro Ala Gly Pro Asn Gly Phe Ala Gly 1962 1971 1980 COG GCT GGT GCT GGT CAA COG GGT GCT AAA GGA GAA AGA GGA GCC AAA GGG Pro Ala Gly Ala Ala Gly Gln Pro Gly Ala Lys Gly Glu Arg Gly Ala Lys Gly 2016 2025 2034 OCT AAG GGT GAR AAC GGT GTT GTT GGT OOC ACA GGC OOC GTT GGA GCT GCT GGC Pro Lys Gly Glu Asn Gly Val Val Gly Pro Thr Gly Pro Val Gly Ala Ala Gly 2079 2088 2097 CON NNN GGT CCA AAT GGT CCC CCC GGT CCT GCT GGA AGT CGT GGT GAT GGA GGC

FIG. 49C

Pro Xxx Gly	Pro As	sn Gly	Pro	Pro	 Gly	 Pro	—- Ala	Gly	 Ser	arg	Gly	 Asp	Gly Gly
2115 CCC CCT GGT	ATG AC		TTC								2151 GGT	<u></u>	2160 CCA GGA
Pro Pro Gly	Met Tr	ur Gly	Phe	Pro	Gly	Ala	Ala	Gly	Arg	Thr	Gly	Pro	Pro Gly
2169		2178			2187			2196			2205		2214
CCC TCT CGT	OT TTA	er ee	CT.	∞ T	GGT	∞	CT	CCI	CCT	GCT	GGG	AAA	CYY CCC
Pro Ser Gly	Ile Se	er Gly	Pro	Pro	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Lys	Glu Gly
222		2222			241			1250				_	2260
CTT CGT GGA	α α	2232 SA GGV	GAÇ		2241 GGA			2250 CCC	CGY		2259 GGA	64A	2268 GIA GEA
Leu Arg Gly	Pro A	rd GTA	W25	GIII	GTA	PIO	ALG	GIÀ	Arg	PIO	GIÀ	GIU	Agr Grå
2277 GCA CCG GGT	~~ ~	2286	رملت		2295	ca c		2304	~~		2313	C) C	2322
							~ -						
Ala Pro Gly	Pro P	ro Gly	Phe	Ala	Gly	Glu	Lys	Gly	Pro	Ser	Gly	Glu	Ala Gly
2331					2349			2358			2367		2376
ACT GCT GGA	α	CT GGC	ACT	CCA	GGT	œ	CAG	GGT	CTT	CTT	GGT	CCI	CCT GGT
Thr Ala Gly	Pro Pr	ro Gly	The	Pro	Gly	Pro	CIV ——	Gly	Leu	Leu	Gly	Ala	Pro Gly
_								2412			2421		2430
2395 ATT CTG GGT	crc a	2394 CT GGC	TCG		2403 GGT	GAA			СТА			GTT	
Ile Leu Gly	Leu P	to GTĀ	Ser	Aig	GIŞ	CIU	Arg	стĀ	Leu	PIO	GLY	V	Ma Giy
2439	C12 ~	2448	ССТ	CT-17	2457	እ ጥጥ	~~~	2466	~~		2475	œ	2484
GCT GTG GGT		CT GGT	CCT	CTT	GGC 	TTA	<u>~~</u>	<u></u>	$\frac{\infty}{1}$	CCT	GGG 		cor cor
2439 GCT GTG GGT Ala Val Gly		CT GGT	CCT	CTT	GGC 	TTA	<u>~~</u>	<u></u>	$\frac{\infty}{1}$	CCT	GGG 		cor cor
GCT GTG GGT Ala Val Gly	Glu P	CT CGT ro Gly 2502	CCT Pro	Leu	GGC Gly 2511	ATT —— Lle	Ala	GGC Gly 2520	CCT Pro	CCT Pro	GGG Gly 2529	Ala	CGT GGT Arg Gly 2538
GCT GTG GGT Ala Val Gly	Glu P	CT CGT ro Gly 2502	CCT Pro	Leu	GGC Gly 2511	ATT —— Lle	Ala	GGC Gly 2520	CCT Pro	CCT Pro	GGG Gly 2529	Ala	CGT GGT Arg Gly 2538
CCT CCT CGT 2493 CCT CCT CGT	Glu P	cr cgr ro Gly 2502 rg cgr	CCT Pro	CCT	GGC Gly 2511 GGA	ATT Ile	Ala AAC	GGC Gly 2520 GGT	Pro	Pro	GGG Gly 2529 GGT	Ala GAA	CGT GGT Arg Gly 2538 GCT GGT
CCT CCT GGT Ala Val Gly 2493 CCT CCT GGT Pro Pro Gly	Glu P	cr ccr ro Gly 2502 rc ccr al Gly	CCT Pro	CCT Leu CCT Pro	GGC Gly 2511 GGA Gly	Ile GIC Val	Ala AAC Asn	GGC Gly 2520 GGT Gly	Pro GCT Ala	Pro	GGG Gly 2529 GGT Gly	Ala GAA	CGT GGT Arg Gly 2538 GCT GGT Ala Gly
CCT GTG GGT Ala Val Gly 2493 CCT CCT GGT Pro Pro Gly	GCT GCT Ala V	cr ccr ro Gly 2502 rc ccr rd Cfr ral Gly	CCT Pro AGT Ser	CCT CCT Pro	GGC Gly 2511 GGA Gly 2565	Ile GIC Val	Ala AAC Asn	GGC Gly 2520 GGT Gly 2574	Pro GCT Ala	CCT Pro CCT Pro	GGG Gly 2529 GGT Gly 2583	Ala GAA Glu	CGT GGT Arg Gly 2538 GCT GGT Ala Gly 2592
CCT GTG GGT Ala Val Gly 2493 CCT CCT GGT Pro Pro Gly 2547 CGT GAT GGC	GCT GALA V	2502 TG GGT TG GGT TG GGT TG GGT TG GGG TG GGG	Pro AGT Ser	CCT Leu CCT Pro	GGC Gly 2511 GGA Gly 2565 GGT	Ile GIC Val	AAC ASn	GGC Gly 2520 GGT Gly 2574 GGT	CCT Pro GCT Ala	CCT Pro CCT Pro	GGG Gly 2529 GGT Gly 2583 GGT	GAA GLu CAA	CGT GGT Arg Gly 2538 CCT GGT Ala Gly 2592 CC GGA
CCT GTG GGT Ala Val Gly 2493 CCT CCT GGT Pro Pro Gly	GCT GALA V	2502 TO GIY 2502 TO GGT 	CCT Pro AGT Ser AAC Asn	CCT Leu CCT Pro GAT Asp	GGC Gly 2511 GGA Gly 2565 GGT Gly	GIC Val	AAC Asn CCA Pro	GGC Gly 2520 GGT Gly 2574 GGT Gly	CCT Pro GCT Ala CCC Arg	Pro CCT Pro GAT Asp	GGG Gly 2529 GGT Gly 2583 GGT Gly	GAA Glu CAA Gln	2538 GCT GGT Arg Gly 2538 GCT GGT Ala Gly 2592 CCC GGA Pro Gly
CCT GTG GGT Ala Val Gly 2493 CCT CCT GGT Pro Pro Gly 2547 CGT GAT GGC Arg Asp Gly	GCT G Ala V AAC O Asn P	cr ccr ro Gly 2502 rc ccr ral Gly 2556 cr ccs ro Gly 2610	AGT Ser AAC	CCT Leu CCT Pro GAT Asp	GGC Gly 2511 GGA Gly 2565 GGT Gly 2619	GIC Val	AAC Asn	GGC Gly 2520 GGT Gly 2574 GGT Gly 2628	CCT Pro GCT Ala CCC Arg	Pro CCT Pro Asp	GGG Gly 2529 GGT Gly 2583 GGT Gly 2637	GAA Glu CAA Gln	CGT GGT Arg Gly 2538 CCT GGT Ala Gly 2592 CCC GGA Pro Gly 2646
CCT GTG GGT Ala Val Gly 2493 CCT CCT GGT Pro Pro Gly 2547 CGT GAT GGC Arg Asp Gly 2601 CAC AAG GGA	GCT GAALA VAAAC OASN P.	2502 TO GLY 2502 TG GGT	AGT Ser AAC ASn TAC	CCT Pro GAT Asp	GGC Gly 2511 GGA Gly 2565 GT Gly 2619 GGC	ATT Lie GTC Val CCC Pro	Ala AAC Asn CCA Pro	GGC Gly 2520 GGT Gly 2574 GGT Gly 2628 GGT	CCT Pro CCT Ala CCC Arg	CCT Pro CCT Pro GAT Asp	2529 GGT Gly 2583 GGT GGT GGT GGT	GAA Glu CAA Gln	2538 CCT GGT Ala Gly 2592 CCC GGA CGT Pro Gly 2646 CCA GGT
CCT GTG GGT Ala Val Gly 2493 CCT CCT GGT Pro Pro Gly 2547 CGT GAT GGC Arg Asp Gly	GCT GAALA VAAAC OASN P.	2502 TO GLY 2502 TG GGT	AGT Ser AAC ASn TAC	CCT Pro GAT Asp	GGC Gly 2511 GGA Gly 2565 GT Gly 2619 GGC	ATT Lie GTC Val CCC Pro	Ala AAC Asn CCA Pro	GGC Gly 2520 GGT Gly 2574 GGT Gly 2628 GGT	CCT Pro CCT Ala CCC Arg	CCT Pro CCT Pro GAT Asp	2529 GGT Gly 2583 GGT GGT GGT GGT	GAA Glu CAA Gln	2538 CCT GGT Ala Gly 2592 CCC GGA CGT Pro Gly 2646 CCA GGT
CCT GTG GGT Ala Val Gly 2493 CCT CCT GGT Pro Pro Gly 2547 CGT GAT GGC Arg Asp Gly 2601 CAC AAG GGA His Lys Gly	GCT GCT GAS OGAS OGAS OGAS OGAS OGAS OGAS OGAS	cT GGT ro Gly 2502 TG GGT ral Gly 2556 cT GGS ro Gly 2610 cc GGT rg Gly 2664	AGT AAC AAC TAC TYr	CCT Leu CCT Pro GAT Asp	GCC Gly 2511 GLy 2565 Gly 2619 GCC Gly 2673	GTC Val	AAC ASn Pro	GGC GIy 2520 GGT GIy 2574 GGT GIy 2628 GGT GIy 2628 GGT GIy 2682	CCT Pro GCT Ala CGC Arg	CCT Pro CCT Pro GAT Asp	GGG GT	GAA Glu CAA Gln GCT Ala	2538 CCT GGT Arg Gly 2538 CCT GGT Ala Gly 2592 CCC GGA Pro Gly 2646 CCA GGT Ala Gly 2700
CCT GTG GGT Ala Val Gly 2493 CCT CCT GGT Pro Pro Gly 2547 CGT GAT GGC Arg Asp Gly 2601 CAC AAG GGA Ris Lys Gly	GCT GCT GAS OGAS OGAS OGAS OGAS OGAS OGAS OGAS	cT GGT ro Gly 2502 TG GGT ral Gly 2556 cT GGS ro Gly 2610 cc GGT rg Gly 2664	AGT AAC AAC TAC TYr	CCT Leu CCT Pro GAT Asp	GCC Gly 2511 GLy 2565 Gly 2619 GCC Gly 2673	GTC Val	AAC ASn Pro	GGC GIy 2520 GGT GIy 2574 GGT GIy 2628 GGT GIy 2628 GGT GIy 2682	CCT Pro GCT Ala CGC Arg	CCT Pro CCT Pro GAT Asp	GGG GT	GAA Glu CAA Gln GCT Ala	2538 CCT GGT Arg Gly 2538 CCT GGT Ala Gly 2592 CCC GGA Pro Gly 2646 CCA GGT Ala Gly 2700
CCT GTG GGT Ala Val Gly 2493 CCT CCT GGT Pro Pro Gly 2547 CGT GAT GGC Arg Asp Gly 2601 CAC AAG GGA His Lys Gly	GCT GAS O	cT GGT ro Gly 2502 TG GGT al Gly 2556 cT GGG ro Gly 2610 CC GGT rg Gly 2664 AT GGC	AGT	CCT Pro GAT Asp CCT Pro	GGC GIy 2511 GGA GIy 2565 GGY 6Iy 2619 GGY 2673 GGY 2673	ATT Ile GTC Val CCC Pro AAT Asn	ALA ASR CCA ASR Pro ATT Lle	GGC Gly 2574 GGT Gly 2628 GGT Gly 2628 GGT Gly 26682 GGT GLY 26682 GGT GLY	CCT Pro CCT Ala CCC Arg CCC Arg	CCT Pro CCT Pro GAT Asp CTT Val	GGG Gly 2529 GGT Gly 2583 GGT Gly 2637 GGT Gly 2691 GGA	GAA Glu CAA Gln GCT Ala	2538 CCT GGT Arg Gly 2538 CCT GGT Ala Gly 2592 CCC GGA
CCT GTG GGT Ala Val Gly 2493 CCT CCT GGT Pro Pro Gly 2547 CGT GAT GGC Arg Asp Gly 2601 CAC AAG GGA His Lys Gly 2655 GCA CCT GGT Ala Pro Gly	GCT GAS O	2502 TO GIY 2502 TG GGT Tal Gly 2556 CT GGG TO GIY 2610 CC GGT TG GIY 2664 AT GGC LIS GIY	AGT	CCT Leu CCT Pro GAT Asp CCT Pro GTG	GGC GIy 2511 GGA GIy 2565 GGY 6Iy 2619 GGY 2673 GGY 2673	ATT LIE GTC Val CCC Pro AAT Asn CCT Pro	AAC ASN CCA Pro ATT Ile	GGC Gly 2574 GGT Gly 2628 GGT Gly 2628 GGT Gly 26682 GGT GLY 26682 GGT GLY	CCT Pro CCT Ala CCC Arg CCC Pro	CCT Pro CCT Asp GAT Val CAT	GGG Gly 2529 GGT Gly 2583 GGT Gly 2637 GGT Gly 2691 GGA	GAA Glu CAA Gln GCT Ala	2538 CCT GGT Arg Gly 2538 CCT GGT Ala Gly 2592 CCC GGA
CCT GTG GGT Ala Val Gly 2493 CCT CCT GGT Pro Pro Gly 2547 CGT GAT GGC Arg Asp Gly 2601 CAC AAG GGA His Lys Gly 2655 GCA CCT GGT	GCT GAS OGLU A	2502 TG GGT TG GGT TG GGT TG GGT TG GGY TG G	AGT AACT ASN TAC CCC Pro	CCT Pro GAT Asp CCT Pro CTG Val	GGC Gly 2511 GGA GGI GGI GGI GGG GGG GGG GGG GGG GGG	ATT Lie GIC Val CCC Pro	AAC AAC AS AS Pro ATT Ile GCT Ala	GGC GIy 2520 GGIY 2574 GGIY 2574 GGIY 2628 GGIY GIY 26882 GGIY GIY 21736	CCT Pro GCT Ala CCC Arg Pro Pro	CCT Pro CCT Pro GAT Asp CAT CAT	GGG GIy 25299 GGT GIy 2583 GGT GIy 2637 GGT GIy 2691 GGA GIy 2745	GAA GLu CAA GLI GLI GLI AL2 AAC ASn	2538 CCT GGT Arg Gly 2538 CCT GGT Ala Gly 2592 CCC GGA Pro Gly 2646 CCA GGT Ala Gly 2700 CGT GGT Arg Gly 2754
CCT GTG GGT Ala Val Gly 2493 CCT CCT GGT Pro Pro Gly 2547 CGT GAT GGC Arg Asp Gly 2601 CAC AAG GGA His Lys Gly 2655 GCA CCT GGT Ala Pro Gly 2709 GAA ACT GGT	GAG O GLU A CCT C Pro H	2502 TG GGT	CCT Pro AGT Ser AAC Asn TAC Tyr CCC Pro	CCT Pro GAT Asp CCT Pro GTG GTG GTG GTT	GSC GIy 2511 GGA GLY 2565 GGT Gly 2619 GGC GLY 2727 GGY 2727 GGY	ATT TILE GTC Val CCC Pro AAT TAN ASI CCT Pro CCT	AAC ASn CCA ATT Ile	GGC GIY 2520 GGT GIY 2574 GGT GIY 2628 GGT GIY 2682 GGC GGY 2736 GGY	CCT Pro CCT Ala CCC Ary CCC Pro AAA CCC CCT CCT CCT CCT CCT CCT	CCT Pro CCT Pro CAT Asp CAT CAT CAT GIT CAT	GGG GIy 2529 GGT GIY 2583 GGT GIY 2637 GIY 2691 GGA GGA GGY 2745 GGY 2745	GAA GLu CAA GLn GCT Ala AAC Asn	2538 CCT GGT ALG GLY 2592 CCC GGA Pro GLY 2646 GCA GGT Ala GLY 2700 CGT GGT ATG GLY 2754 ACG GGT
CCT GTG GGT Ala Val Gly 2493 CCT CCT GGT Pro Pro Gly 2547 CGT GAT GGC Arg Asp Gly 2601 CAC AAG GGA His Lys Gly 2655 GCA CCT GGT Ala Pro Gly	GCT GAG OGAS PROBLEM AND CCT CCT CCT CCT CCT CCT CCT CCT CCT CC	2502 TG GGT	AGT	CCT Leu CCT Pro GAT Asp CCT Pro GTG GTG Val	GSC GIy 2511 GGA GLY 2565 GGT Gly 2619 GGC GLY 2727 GGY 2727 GGY	ATT Ile GTC Val CCC Pro AAT Asn CCT Pro CCT Pro	AAC ASN CCA ATT Ile GCT Ala	GGC GIY 2520 GGT GIY 2574 GGT GIY 2628 GGT GIY 2682 GGC GGY 2736 GGY	CCT Pro GCT Ala CCC Arg Pro AAA CCC ARG CCC AAA AAA AAA AAA	CCT Pro CCT Pro CCT Asp CAT Asp Val CAT CAT Val	GGG GIy 2529 GGT GIY 2583 GGT GIY 2637 GIY 2691 GGA GGA GGY 2745 GGY 2745	GAA GIU CAA GII GII Ala AC ASI CCA Pro	2538 CCT GGT ALG GLY 2592 CCC GGA Pro GLY 2646 GCA GGT Ala GLY 2700 CGT GGT ATG GLY 2754 ACG GGT

FIG. 49D

Pro Ser Gly Pro Gln Gly Ile Arg Gly Asp Lys Gly Glu Pro Gly Glu Lys Gly 2817 2826 2835 2844 2853 2862 CCC AGA GGT CTT CCT GGC TTC AAG GGA CAC AAT GGA TTG CAA GGT CTG CCT GGT Pro Arg Gly Leu Pro Gly Phe Lys Gly His Asn Gly Leu Gln Gly Leu Pro Gly ATC GCT GGT CAC CAT GGT GAT CAA GGT CCT CCT GGC TCC GGT GGT CCT GGT Ile Ala Gly His His Gly Asp Gln Gly Ala Pro Gly Ser Val Gly Pro Ala Gly CCT AGG GGC CCT GCT GGT CCT TCT GGC CCT GCT G	α	AGT	GGC	CCA	CAA	GGC	ATT	CCT	GGC	Cat	AAG	GGA	GAG	∞	GGT	GAA	aag	GGG
CCC AGA GGT CTT CCT GGC TTC AAG GGA CAC AAT GGA TTG CAA GGT CTG CCT GGT Pro Arg Gly Leu Pro Gly Phe Lys Gly His Asn Gly Leu Gln Gly Leu Pro Gly 2871 2880 2889 2898 2907 2916 ATC GCT GGT CAC CAT GGT GAT CAA GGT GCT CCT GGC TCC GTG GGT CCT GCT GGT Ile Ala Gly His His Gly Asp Gln Gly Ala Pro Gly Ser Val Gly Pro Ala Gly 2925 2934 2943 2952 2961 2970 CCT AGG GGC CCT GCT GGT CCT TCT GGC CCT GCT G	Pro	Ser	Gly	Pro	GTV	Gly	Ile	Arg	Gly	Asp	Lys	Gly	Glu	Pro	Gly	Glu	Lys	Gly
CCC AGA GGT CTT CCT GGC TTC AAG GGA CAC AAT GGA TTG CAA GGT CTG CCT GGT Pro Arg Gly Leu Pro Gly Phe Lys Gly His Asn Gly Leu Gln Gly Leu Pro Gly 2871 2880 2889 2898 2907 2916 ATC GCT GGT CAC CAT GGT GAT CAA GGT GCT CCT GGC TCC GTG GGT CCT GCT GGT Ile Ala Gly His His Gly Asp Gln Gly Ala Pro Gly Ser Val Gly Pro Ala Gly 2925 2934 2943 2952 2961 2970 CCT AGG GGC CCT GCT GGT CCT TCT GGC CCT GCT G			2817		:	2826		:	2835		:	2844		;	2853		:	2862
2871 2880 2889 2898 2907 2916 ATC GCT GGT CAC CAT GGT GAT CAA GGT GCT CCT GGC TCC GGG GGT CCT GCT GGT Ile Ala Gly His His Gly Asp Gln Gly Ala Pro Gly Ser Val Gly Pro Ala Gly 2925 2934 2943 2952 2961 2970 CCT AGG GGC CCT GCT GGT CCT TCT GGC CCT GGT AGA AAA GAT GGT CGC ACT GGA Pro Arg Gly Pro Ala Gly Pro Ser Gly Pro Ala Gly Lys Asp Gly Arg Thr Gly 2979 2988 2997 3006 3015 3024 CAT CCT GGT AGG GTT GGA CCT GCT GGC ATT CGA GGC CCT CAG GGT CAC CAA GGC His Pro Gly Thr Val Gly Pro Ala Gly Ile Arg Gly Pro Gln Gly His Gln Gly 3033 3042 3051 3060 3069 3078 CCT GCT GGC CCC CCT GGT CCC CCT GGC CCT CTT GGA CCT CTA GGT GTA AGC GGT Pro Ala Gly Pro Pro Gly Pro Pro Gly Pro Leu Gly Pro Leu Gly Val Ser Gly 3087 3096 3105 3114	∞															CIG		
2871 2880 2889 2898 2907 2916 ATC GCT GGT CAC CAT GGT GAT CAA GGT GCT CCT GGC TCC GGG GGT CCT GCT GGT Ile Ala Gly His His Gly Asp Gln Gly Ala Pro Gly Ser Val Gly Pro Ala Gly 2925 2934 2943 2952 2961 2970 CCT AGG GGC CCT GCT GGT CCT TCT GGC CCT GGT AGA AAA GAT GGT CGC ACT GGA Pro Arg Gly Pro Ala Gly Pro Ser Gly Pro Ala Gly Lys Asp Gly Arg Thr Gly 2979 2988 2997 3006 3015 3024 CAT CCT GGT AGG GTT GGA CCT GCT GGC ATT CGA GGC CCT CAG GGT CAC CAA GGC His Pro Gly Thr Val Gly Pro Ala Gly Ile Arg Gly Pro Gln Gly His Gln Gly 3033 3042 3051 3060 3069 3078 CCT GCT GGC CCC CCT GGT CCC CCT GGC CCT CTT GGA CCT CTA GGT GTA AGC GGT Pro Ala Gly Pro Pro Gly Pro Pro Gly Pro Leu Gly Pro Leu Gly Val Ser Gly 3087 3096 3105 3114	 B	~	Cly	Tou	Pro	Gly	Dhe	Tare	Gly			Gly	Len	Gln	Gly			~~~
ATC GCT GGT CAC CAT GGT GAT CAA GGT GCT CCT GCC TCC GGG GGT CCT GCT GGT Ile Ala Gly His His Gly Asp Gln Gly Ala Pro Gly Ser Val Gly Pro Ala Gly 2925 2934 2943 2952 2961 2970 CCT AGG GGC CCT GCT GGT CCT TCT GGC CCT GCT G	510	λtg	GIY	Deu	110	GLY		uy 3	OLY	.43	٠	ary	L	GL.	GLY	LEU	PLO	GTÅ
Ile Ala Gly His His Gly Asp Gln Gly Ala Pro Gly Ser Val Gly Pro Ala Gly 2925 2934 2943 2952 2961 2970 CCT AGG GGC CCT GCT GGT CCT TCT GGC CCT GCT G																		
2925 2934 2943 2952 2961 2970 CCT AGG GGC CCT GGT GGT CCT TCT GGC CCT GGA AAA GAT GGT CGC ACT GGA Pro Arg Gly Pro Ala Gly Pro Ser Gly Pro Ala Gly Lys Asp Gly Arg Thr Gly 2979 2988 2997 3006 3015 3024 CAT CCT GGT ACG GTT GGA CCT GCT GGC ATT CGA GGC CCT CAG GGT CAC CAA GGC His Pro Gly Thr Val Gly Pro Ala Gly Ile Arg Gly Pro Gln Gly His Gln Gly 3033 3042 3051 3060 3069 3078 CCT GCT GGC CCC CCT GGT CCC CCT GGC CCT CTT GGA CCT CTA GGT GTA AGC GGT Pro Ala Gly Pro Pro Gly Pro Pro Gly Pro Leu Gly Pro Leu Gly Val Ser Gly 3087 3096 3105 3114	ATC	GCT	GGT	CAC	CAT	GGT	GAT	CAA	GGT	CCT	œ	GCC	TCC	GIG	GGT	œ	CCI	GGT
2925 2934 2943 2952 2961 2970 CCT AGG GGC CCT GGT GGT CCT TCT GGC CCT GGA AAA GAT GGT CGC ACT GGA Pro Arg Gly Pro Ala Gly Pro Ser Gly Pro Ala Gly Lys Asp Gly Arg Thr Gly 2979 2988 2997 3006 3015 3024 CAT CCT GGT ACG GTT GGA CCT GCT GGC ATT CGA GGC CCT CAG GGT CAC CAA GGC His Pro Gly Thr Val Gly Pro Ala Gly Ile Arg Gly Pro Gln Gly His Gln Gly 3033 3042 3051 3060 3069 3078 CCT GCT GGC CCC CCT GGT CCC CCT GGC CCT CTT GGA CCT CTA GGT GTA AGC GGT Pro Ala Gly Pro Pro Gly Pro Pro Gly Pro Leu Gly Pro Leu Gly Val Ser Gly 3087 3096 3105 3114	T10	212	Glu	ui e	Hie	Gly	Asp	Gln	Gly	Δla	Pro	Gly	Ser	Val	Gly	Dro	N12	Clu
CCT AGG GGC CCT GGT GGT CCT TCT GGC CCT GGA AAA GAT GGT CGC ACT GGA Pro Arg Gly Pro Ala Gly Pro Ser Gly Pro Ala Gly Lys Asp Gly Arg Thr Gly 2979 2988 2997 3006 3015 3024 CAT CCT GGT AGG GTT GGA CCT GCT GGC ATT GGA GGC CCT CAG GGT CAC CAA GGC His Pro Gly Thr Val Gly Pro Ala Gly Ile Arg Gly Pro Gln Gly His Gln Gly 3033 3042 3051 3060 3069 3078 CCT GCT GGC CCC CCT GGT CCC CCT GGC CCT CTT GGA CCT CTA GGT GTA AGC GGT Pro Ala Gly Pro Pro Gly Pro Pro Gly Pro Leu Gly Pro Leu Gly Val Ser Gly 3087 3096 3105 3114	116	Λια	GIŞ	143		OLy	ΑÇA	٠	OL,	744		OLY	JCL	٧	Gry	110	Ala	GIĀ
Pro Arg Gly Pro Ala Gly Pro Ser Gly Pro Ala Gly Lys Asp Gly Arg Thr Gly 2979 2988 2997 3006 3015 3024 CAT CCT GGT ACG GTT GGA CCT GCT GGC ATT CGA GGC CCT CAG GGT CAC CAA GGC His Pro Gly Thr Val Gly Pro Ala Gly Ile Arg Gly Pro Gln Gly His Gln Gly 3033 3042 3051 3060 3069 3078 CCT GCT GGC CCC CCT GGT CCC CCT GGC CCT CTT GGA CCT CTA GGT GTA ACC GGT Pro Ala Gly Pro Pro Gly Pro Pro Gly Pro Leu Gly Pro Leu Gly Val Ser Gly 3087 3096 3105 3114																		
2979 2988 2997 3006 3015 3024 CAT CCT GGT ACG GTT GGA CCT GCT GGC ATT CGA GGC CCT CAG GGT CAC CAA GGC His Pro Gly Thr Val Gly Pro Ala Gly Ile Arg Gly Pro Gln Gly His Gln Gly 3033 3042 3051 3060 3069 3078 CCT GCT GGC CCC CGT GGT CCC CCT GGC CCT CTT GGA CCT CTA GGT GTA AGC GGT Pro Ala Gly Pro Pro Gly Pro Pro Gly Pro Leu Gly Pro Leu Gly Val Ser Gly 3087 3096 3105 3114	CCT	AGG	GGC	α	CCT	GGT	∞ r	TCT	GGC	α	CT	CC4	AAA	GAT	GCI	α c	ACT	GGA
2979 2988 2997 3006 3015 3024 CAT CCT GGT ACG GTT GGA CCT GCT GGC ATT CGA GGC CCT CAG GGT CAC CAA GGC His Pro Gly Thr Val Gly Pro Ala Gly Ile Arg Gly Pro Gln Gly His Gln Gly 3033 3042 3051 3060 3069 3078 CCT GCT GGC CCC CGT GGT CCC CCT GGC CCT CTT GGA CCT CTA GGT GTA AGC GGT Pro Ala Gly Pro Pro Gly Pro Pro Gly Pro Leu Gly Pro Leu Gly Val Ser Gly 3087 3096 3105 3114	D	~~~		2	A1 a	Gly	D-0	502	Glu	Pro	A12	Gly	Tue	200		 N		
CAT CCT GGT ACG GTT GGA CCT GCT GGC ATT CGA GGC CCT CAG GGT CAC CAA GGC His Pro Gly Thr Val Gly Pro Ala Gly Ile Arg Gly Pro Gln Gly His Gln Gly 3033 3042 3051 3060 3069 3078 CCT GCT GGC CCC CCT GGT CCC CCT GGC CCT CTT GGA CCT CTA GGT GTA AGC GGT Pro Ala Gly Pro Pro Gly Pro Pro Gly Pro Leu Gly Pro Leu Gly Val Ser Gly 3087 3096 3105 3114	PIO	ALG	Gry	210	٠'na	GLY	FLO	Jez	Gry	110	Ata	Gry	плэ	wsb	GTÅ	Atg	me	GIÀ
His Pro Gly Thr Val Gly Pro Ala Gly Ile Arg Gly Pro Gln Gly His Gln Gly 3033 3042 3051 3060 3069 3078 CCT GCT GGC CCC CCT GGT CCC CCT GGC CCT CTT GGA CCT CTA GGT GTA AGC GGT Pro Ala Gly Pro Pro Gly Pro Pro Gly Pro Leu Gly Pro Leu Gly Val Ser Gly 3087 3096 3105 3114																		
3033 3042 3051 3060 3069 3078 CCT GCT GGC CCC GGT CCC CCT GGC CCT CTT GGA CCT CTA GGT GTA AGC GGT Pro Ala Gly Pro Pro Gly Pro Pro Gly Pro Leu Gly Pro Leu Gly Val Ser Gly 3087 3096 3105 3114	CAT	α	GGT	ACG	GTT	GGA	α	CCT	GCC	TTA	ŒA	ccc	α	CAG	GGT	CAC	CAA	GGC
3033 3042 3051 3060 3069 3078 CCT GCT GGC CCC GGT CCC CCT GGC CCT CTT GGA CCT CTA GGT GTA AGC GGT Pro Ala Gly Pro Pro Gly Pro Pro Gly Pro Leu Gly Pro Leu Gly Val Ser Gly 3087 3096 3105 3114	His	Pro	Gly			_												
Pro Ala Gly Pro Pro Gly Pro Pro Gly Pro Leu Gly Pro Leu Gly Val Ser Gly 3087 3096 3114				Thr	Val	Glv	Pro	Δla	Glv	Tle	A	Gly	Pm	Gla	Glu	wie.	Cl=	G)
Pro Ala Gly Pro Pro Gly Pro Pro Gly Pro Leu Gly Pro Leu Gly Val Ser Gly 3087 3096 3105 3114			GLY	Thr											_			GJA
3087 3096 3105 3114		:	3033		:	3042		:	3051		3	3060		3	3069		3	078
3087 3096 3105 3114	CCT	:	3033		:	3042		:	3051		3	3060		3	3069		3	078
3087 3096 3105 3114 GGT GGT TAT GAC TTT GGT TAC GAT GGA GAC TTC TAC AGG GCT 3'		GCT	3033 GGC	<u></u>	<u>ст</u>	3042 GGT	<u></u>	ст 	3051 GGC	СТ ——	CIT	9060 9060	œr ——	CTA	3069 GGT	GTA	3 AGC	078 GGT
GGT GGT TAT GAC TIT GGT TAC GAT GGA GAC TIC TAC AGG GCT 3'		GCT Ala	Gly	œc Pro	CCT Pro	GGT GGT Gly	œ Pro	CT Pro	Gly	CCT —— Pro	CTT —— Leu	9060 9060	œr ——	CTA	3069 GGT	GTA	3 AGC	078 GGT
	Pro	GCT Ala	3033 GGC Gly 3087	œc Pro	CT Pro	3042 GGT Gly 3096	œ Pro	CT Pro	Gly B105	CCT Pro	CTT —— Leu	Gly 6114	CCT Pro	CTA Leu	Gly GGT	GTA	3 AGC	078 GGT
Gly Gly Tyr Asp Phe Gly Tyr Asp Gly Asp Phe Tyr Arg Ala	Pro	GCT Ala	3033 GGC Gly 3087	œc Pro	CT Pro	3042 GGT Gly 3096	œ Pro	CT Pro	Gly B105	CCT Pro	CTT —— Leu	Gly 6114	CCT Pro	CTA Leu	Gly GGT	GTA	3 AGC	078 GGT

FIG. 49E

36 45 54 GGT CCG ATG GGT		27 CTG	GGC	gta	18 GGC	AAA	GGT	9 GAC	TAC	CAG	51
Gly Pro Met Gly Leu Met Gly											
90 99 108 GGT CCG CAG GGC TTC CAA GGT Gly Pro Gln Gly Phe Gln Gly	CCG CCG GGT										
144 153 162 GGT CCG GCG GGT GCT CGC GGT	144 CAG ACG GGT	135 GGT	CCG	GAA	126 GGC	ccc	GAA	117 GGT	GCG	ccc	
198 207 216 GGT CAC CCG GGT AAG CCA GGC	198	189			180			171			
Gly His Pro Gly Lys Pro Gly											
252 261 270 A GGT GCG CGT GGT TTC CCG GGC	CCG CAA GG						GAA				٠
306 315 324 GGT CAC AAC GGT CTG GAC GGT	306	297			288			279		_	
Gly His Asn Gly Leu Asp Gly	Ile Arg Gly	Gly	Lys	Phe	Gly	Pro	Leu	Gly	Pro	Thr	
360 369 378 GGC GAA CCG GGT GCC CCA GGC Gly Glu Pro Gly Ala Pro Gly	GTC AAA GGC						CAA				
414 423 432 GGT CTG CCG GGT GAA CGC GGC	414 GCG CGT GGT	405 GGT	ACT	CAG	396 GGC	ccc	ACG	387 GGT	AAC	GAA	
Gly Leu Pro Gly Glu Arg Gly 468 477 486 GGC AGC GAT GGC TCC GTC GGT	468	459			450			441			
Gly Ser Asp Gly Ser Val Gly	Ala Arg Gly	Gly	Ala	Pro	Gly	Pro	Ala	Gly	.Val	Arg	
522 531 540 GGC CCT CCG GGT TTC CCG GGT	roc got ggo										
Gly Pro Pro Gly Phe Pro Gly	Ser Ala Gly	Gly	Ile	Pro		Ala	Pro		Val	Pro •	
GGC AAC GCA GGC CCG GCT GGT		567 GGC	ATC	GAG		AAG	CCG		CCG	GCG	
Gly Asn Ala Gly Pro Ala Gly	Ala Val Gly	Gly	Ile	Glu	Gly	Lys	Pro	Gly	Pro	Ala	
Gly Leu Ser Gly Pro Val Cly	CTG CCG GGT										
GGT CTG CCG GGT GAA CGC Gly Leu Pro Gly Glu Arg 468 477 GGC AGC GAT GGC TCC GTC Gly Ser Asp Gly Ser Val 522 531 GGC CCT CCG GGT TTC CCG Gly Pro Pro Gly Phe Pro 576 GGC AAC GCA GGC CCG GCT Gly Asn Ala Gly Pro Ala 630 639	Ala Arg Gly Ala Arg Gly GCG CGT GGC Ala Arg Gly Ser Ala Gly GCG GTT GGC Ala Val Gly CTG CCG GGT	GGT Gly 459 GGC Gly 513 GGT GGC Gly 621 GGT	Thr GCT Ala ATT Ile ATC Ile GTC	Gln CCG Pro CCG Pro GAG Glu GAA	GGC Gly 450 GGT 504 GGT Gly 558 GGT Gly 612 GGC	Pro CCG Pro GCG Ala AAG Lys	ACG Thr GCT Ala CCT Pro CCG Pro	GGT Gly 441 GGC Gly 495 GGC Gly 603 GGT Gly 603 GGC	Asn GTT Val, CCG Pro	CCG Pro GCG Ala	

CCA	CĊG	657 GGT	aac	CCG	666 GGC	GCA	AAC	675 GGC-	CTG	ACG	684 GGT	GCA	AAA	693 GGT	GCG	GCT	702 GGC
			Asn														
Pro	Pro	Gly	ASI	PIO	GIA	wra	WOII		٠	****		****	2,0				_
	~~~	711	GTT	CCC	720	GCC	CCG	729 GGC	CTG	CCG	738 GGT	CCG	CGC	747 GGT	ATT	CCG	756 GGT
Leu	Pro	Gly	Val	Ala	Gly	Ala	Pro	Gly	Leu	Pro	Gly	Pro	Arg	GIA	116	Pro	GIA
		765			774			783			792			801	<b>~</b>	~~	810
			GCA														
Pro	Val	Gly	Ala	Ala	Gly	Ala	Thr	Gly	Ala	Arg	Gly	Leu	Val	Gly	Glu	Pro	Gly
		819			828			837			846			855			864
CCC	GCG	CCT	TCT	AAA	GGC	GAA	AGC	GGŢ	AAC	AAA	GGT	GAG	CCG	GGT	TCC	GCG	GGC
Pro	Ala	Gly	Ser	Lys	Gly	Glu	Ser	Gly	Asn	Lys	Gly	Glu	Pro	Gly	Ser	Ala	Gly
		873			882			891			900			909			918
CCG	CAG	GGT	CCG	CCG	GGT	CCG	AGC	GGC	GAA	GAA	GGT	AAA	CGT	GGT	ccc	AAC	GGC
Pro	Gln	Gly	Pro	Pro	Gly	Pro	Ser	Gly	Glu	Glu	Gly	Lys	Arg	Gly	Pro	Asn	Gly
		927		•	936			945			954			963			972
GAG	CCT	GGT	TCC	GCA	GGC	CCT	ccc	CCT	ccc	000	CCT	CTG	CCT	CCC	AGC	œ	CCT
Glu	Ala	Gly	Ser	Ala	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Leu	Arg	Gly	Ser	Pro	Gly
		981			990			999		:	1008		:	1017			1026
AÇC	CGT	GGC	CIG	CCG	GGC	GCG	GAC	GGC	CGT	GCG	GGC	GTG	ATG	GGT	CCG	CCG	GGT
Ser	Arg	Gly	Leu	Pro	Gly	Ala	Asp	Gly	Arg	Ala	Gly	Val	Met	Gly	Pro	Pro	Gly
		1035			1044			1053			1062			1071			1080
TCC	CCT	GGT	GCC	TCT	GGT	CCG	GCT	GGT	GTC	CCI	GGT	CCG	AAT	GGC	GAC	GCG	GGC
Ser	Arg	Gly	Ala	Ser	Gly	Pro	Ala	Gly	Val	Arg	Gly	Pro	Asn	Gly	Asp	Ala	Gly
		1089			1098			1107		:	1116		1	L125		1	L134
CGT	CCG	GGT	GAA	CCG	GGC	CIG	ATG	GGT	ccc	CCI	GGC	CTG	CCCG	GGT	AGC	CCG	GGT
Arg	Pro	Gly	Glu	Pro	Gly	Leu	Met	Gly	Pro	Arg	Gly	Leu	Pro	Gly	Ser	Pro	Gly
	:	1143			1152		:	1161			1170		1	L179		1	L188
AAC	ATT	CCT	CCC	GCG	GGT	AAG	GAG	CCT	CCC	GTA	GGT	CTG	ccc	GGT	ATT		
Asn	Ile	Gly	Pro	Ala	Gly	Lys	Glu	Gly	Pro	Val	Gly	Leu	Pro	Gly	Ile	Asp	Gly
		1197			1206			1215			1224			- L233	•		L242
CGT	CCG	GGT	CCG	ATC	GGC	CCT	GCG	GGC	GCT	CGT	GGC	GAG	CCG	GGT	AAC	ATC	GGT
Arg	Pro	Gly	Pro	Ile	Gly	Pro	Ala	Gly	Ala	Arg	Gly	Glu	Pro	Glv	Asn	Ile	Gly
		1251			1260			1269			1278	·		1287			296
TTT			CCG			CCG			GAC	ccc	GGC	AAG	AAC	GGT	CAT	yyy ,	GGC
Phe	Pro	Gly	Pro	Lys	Gly	Pro	Thr	Gly	Asp	Pro	Gly	Lys	Asn	Gly	asp	Lys	Gly
		1305			1314		:	1323		3	1332		1	341		1	.350
CAT			CIG			GCC			GCA	CCC	CCT	ccc	GAT	CCT	AAC	AAT	GGT
His	Ala	Gly	Leu	Ala	Gly	Ala	Arg	Gly	Ala	Pro	Gly	Pro	Asp	Gly	Asn	Asn	Gly

1359 GCG CAG GGT			1377 CAG GGC					1404 GAA CAG OGT
Ala Gln Gly	Pro Pro	Gly Pro	Gln Gly	Val	Gln Gly	Gly Lys	Gly	Glu Gln Gly
1413 CCG GCA GGC		GGC TTC	1431 CAG GGT		1440 CCG GGT	CCG ACC	1449 GGC	1458 CCG GCT GGT
Pro Ala Gly	Pro Pro	Gly Phe	Gln Gly	Leu	Pro Gly	Pro Sei	Gly	Pro Ala Gly
	AAA CCG		CGT GGC		CAT GGC		GGC	CTG CCG GGT
Glu Val Gly	Lys Pro	Gly Glu	Arg Gly	Leu	His Gly	Glu Phe	Cly	Leu Pro Gly
	CCG CGT		CGC GGC		ccc ccc	GAA TCC		1566 GCG GCA GGT  Ala Ala Gly
	CCG ATT	GGT TCC		CCG	AGC GGC		GGT	1620 CCG GAC GGC Pro Asp Gly
1629 AAC AAA GGC			1647 GTT GGT			ACC GCC		1674 CCG TCT GGT
Asn Lys Gly	Glu Pro	Gly Val	Val Gly	Ala	Val Gly	Thr Ala	Gly	Pro Ser Gly
	CTG CCG			GCC -				1728 GGC AAA GGT Gly Lys Gly
	GAA CCG			GAG				1782 CGT GAC GGT Arg Asp Gly
	GCA CAC			GCA				1836 GCG ACT GGT Ala Thr Gly
	GAA GCT		GCG GGT		GCG GCT			1890 CCT CGC GGT  Pro Arg Gly
	GAA CGC			CCG		CCG AAT		1944 TTT GCT GGC  Phe Ala Gly
				GCG				1998 GCC AAA GGC
2007		2016	2025		2034		2043	Ala Lys Gly 2052
							GGT	OCG GCT GGT Ala Ala Gly
,1		,				-20 10.		u oay

	2	061		2	070	000	ccc 2	2079	~~	CC A	880	MCC.	رين ر	2097			2106
CCG (																	
Pro	Ala	Gly	Pro	Asn	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Ser	Arg	Gly	ASD	GIĀ	GIĀ
	2	115		2	2124			2133	~~	~~	2142	œ		2151	സ്ത		2160
		GGC															
Pro	Pro	Gly	Met	Thr	Gly	Phe	Pro	Gly	Ala	Ala	Gly	Arg	Thr	Gly	Pro	Pro	Gly
		2169		. :	2178			2187			2196			2205			2214
		GGC															
Pro	Ser	Gly	Ile	Ser	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Pro	Ala	Gly	Lys	Glu	Gly
	:	2223		:	2232			2241			2250			2259			2268
CIG	CGT	GGC	CCA	CGC	GGC	GAC	CAG	GGT	CCG	GTG	GGC	CGT	ACC	GGC	GAA	GIC	GGT
Leu	Arg	Gly	Pro	Arg	Gly	Asp	Gln	Gly	Pro	Val	Gly	Arg	Thr	Gly	Glu	Val	Gly
		2277			2286			2295			2304			2313			2322
		GGC					~~-										~~~
Ala	Val	Gly	Pro	Pro	Gly	Phe	Ala	Gly	Glu	Lys	Gly	Pro	Ser	Gly	Glu	Ala	Gly
		2331			2340			2349						2367			2376
		GGC															
Thr	Ala	Gly	Pro	Pro	Gly	Thr	Pro	Gly	Pro	Gln	Gly	Leu	Leu	Gly	Ala	Pro	Gly
		2385			2394			2403			2412	~~~		2421			2430
		GGC															
Ile	Leu	Gly	Leu	Pro	Gly	Ser	Arg	Gly	Glu	Arg	Gly	Leu	Pro	Gly	Val	Ala	Gly
~~		2439 . GGC			2448			2457			2466			2475			2484
	~																
Ala	Val	Gly	Glu	Pro	Gly	Pro	Leu	Gly	Ile	Ala	Gly	Pro	Pro	Gly	Aļa	Arg	Gly
CCC:	~~	2493 GGT			2502			2511			2520	œ		2529	(23.3		2538
Pro	Pro	Gly	Ala	Val	Gly	Ser	Pro	Gly	Val	Asn	Gly	Ala	Pro	Gly	Glu	Ala	Gly
œ	CAC	2547 GGC			2556			2565			2574	CCT		2583	CAG		2592
Arg	ASD	GIY	ASD	Pro	GTA	ASD	ASp	GIĀ	Pro	Pro	GIĀ	Arg	Asp	GIÅ	Gīu	Pro	Gly
CAC	AAA	2601 GGT		CGT	2610 GGC		CCG	2619 GGT			2628 GGT	സ		2637 CGT	സ്ത		2646 GGC
		~															
HIS	гĀЗ	Gly	GIU	Arg	GIŞ	lyr	Pro	GIÅ	ASD	· Ile	GIĀ	Pro	Val	Gly	Ala	Ala	Gly
GCT	CCG	2655 GGT		CAC	2664 GGT		GTA	2673 GGC			2682 GGC	AAA		2691 GGT	AAC		2700 GGT
								·									
wrq	FIO	Gly		uis			val					กังล		_	ASN	_	_
GAA	ACG	2709 GGT		TCC	2718 CCT		GTA	2727 GGT			2736 GGT	CCT	GTT	2745 GGT	CCA	ccc	2754 GGC
																	Gly
	****	-TA			- J-7						7		- 41	GTA	LIO	ruy	OTA

	:	2763			2712			2781			2790			.:799			2808
CCC	TCC	GGC	CCG	CAG	GGT	ATT	CGC	GGT	GAC	AAA	GGC	GAA	CCG	GGC	GAλ	AAA	GGT
		<del></del>															
Pro	Ser	Gly	Pro	Gln	Gly	Ile	Arg	Gly	Asp	Lys	Gly	Glu	Pro	Gly	Glu	Lys	Gly
					2826												2862
CCC	CCT	CCT	CTG	സ്ക	GGC	CIT	AAG	GGC	CAC	AAC	GGT	CTG	CAA	GGT	CTG	CCG	GGT
Pro	Arg	Gly	Leu	Pro	Gly	Leu	Lys	СīЛ	His	Asn	СŢУ	· Leu	Gin	GIY	Leu	Pro	Gly
		2071			2880			2880			2000			מחם כ			2916
»mv															CCG		
Ile	Ala	Gly	His	His	Gly	Asp	Gln	Gly	Ala	Pro	Gly	Ser	Val	Gly	Pro	Ala	Glv
											_			•			2
		2925			2934												2970
CCC	CCT	GGC	CCC	GCT	GGT	CCC	TCT	GGT	ĊCC	GCC	CGT	AAA	GAC	GGC	CGT	ACG	GGC
Pro	Arg	Gly	Pro	Ala	Gly	Pro	Ser	Gly	Pro	Ala	Gly	Lys	Asp	Gly	Arg	Thr	Gly
		2979			2088			2007			3006		-			_	
CAC				CIC	TOOK 2	ന്നു	GCC.	722	יוידע	CCC.	ייבר) מטטנ	~~	C22	0.T.5	CAC	220	3024
											-551		CAA	GGT		CAL	GGT
His	Pro	Gly	Thr	Val	Gly	Pro	Ala	Glv	Ile	Arg	Gly	Pro	Gln	Glv	His	Gln	Cly
																	GIŞ
		3033			3042			3051		:	3060		3	3069		3	078
ccc	GCG	GGT	ccc	CCC	CCT	CCC	$\infty$	CCT	$\infty$	ccc	CCT	CCC	CCC	CCT	GTT	AGC	GGT
PIO	ALA	GIA	Pro	PTO	GTA	PTO	Pro	Gly	Pro	Pro	Gly	Pro	Pro	Gly	Val	Ser	Gly
	-	3087		•	3096			2105		3							
GGC				TTT	GGT	TAT	GAC	1500 1707	CAT	مكلس	መጀመ ትቸቸ	<b>~~</b> π	~~	٠.			
							~~~							2.			
Gly	Gly	Tyr	Asp	Phe	Gly	Tyr	Asp	Gly	Asp	Phe	Tyr	Arg	Ala				

FIG. 50E

 $\overline{\rm EcoR1}$ start 0ligo N1-1 ($lpha_2$) 5-66aattcatggcagtatgatggcagaaggcgtcggccccgggcccaatggccctcatgggcccccgcggccca-3'

BsrFl stop Hind III

3'-ccesecceccecceccaceteccaceccecececececetecaacetecaagetecces

Oligo N1-2 (α_2)

EcoR1 BSRF1 Oligo N1-3 (α_2)

5-GGAATTCGCCGGTGAGCCGGGTGAACCGGGCCAAACGGGTCCGCCAGGTGCACGTGGTCCAGCGGGCCCGCCTGGCAAGGCG-3'

3'-CCGGGCGGACCGTTCCGCCACTTCTACCGGTGGGACCGTTTGGCCCGGGGGCCCACTCGCACCGCATCACATATTCGAACCC-5'

Oligo N1-4 (α_2)

FIG

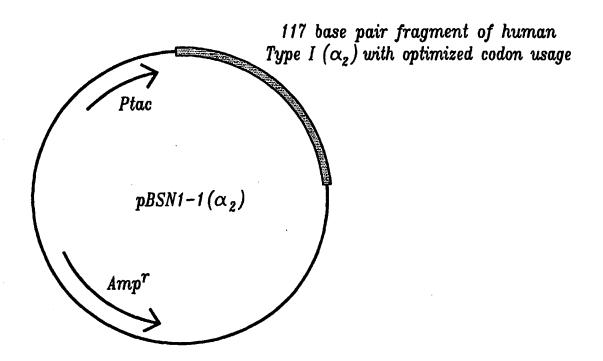


FIG. 52

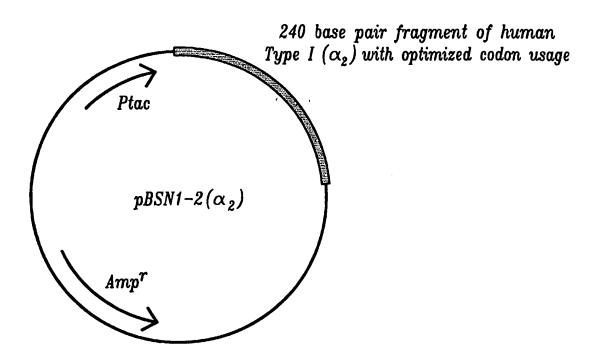


FIG. 53

			9			18			27			36			45			54
5'	CAG	TAT	GAT	GGC	AAA	GCC	GTC	GGC	CIC	œc	∞	GCC	CCA	ATG	GGC	CTC	ATG	GGC
															~			
	Gln	Tyr	Asp	Gly	Lys	Gly	Val	Gly	Leu	Gly	Pro	Gly	Pro	Met	Gly	Leu	Met	Gly
			63			72			81	•	,	•						
	~~~	~~~		~~	~~		~~	~~				90			99			108
	W.	usc	GGC	UA	u.	GGI	GLA	GCT	GGC	GC.	CCA	GGC	œ	CAA	GGT	TIC	CAG	GGC
	Pro	Arg	Gly	Pro	Pro	Gly	ALA	ALA	Gly	Ala	Pro	Gly	Pro	Gln	Gly	Phe	Gln	GJÀ
			117			126			125									
	~~	~~		010	~~				135			144			153			162
	U.I	نك	GGT	CHIL	u	GGT	GAA	u.	GGC	CAA	ACG	GGT	$\infty$	GCA	GGT	CCA	CGT	GGT
				01														
	Pro	ALA	Gly	GTIT	Pro	GTĀ	GLU	Pro	Gly	Gln	Thr	Gly	Pro	Ala	Gly	Ala	Arg	Gly
			171			180			189			198						
	CCA		GGC	ന്നു	(Tr		224	ccc	COT	~ »	~~	~~	~~		207			216
					~				931	(ZW	G-11	ناتانا	CAL	C.I	GGC	AAA	$\infty$	GGC
	Pm	Δla	Glu	Pm	Dro	Cly	Tare	212	<u></u>	Cl	<u> </u>	~						
			Gly		***	GLy.	my 3	ALC.	GIY	GIU	ASD	GrÅ	HLS	Pro	GLY	Lys	Pro	Gly
			225			234												
	$\infty$	$\infty$	<b>GGT</b>	GAG			GTA	GTG										
	Arg	Pro	Gly	Glu	Arg	Gly	Val	Val										
						_												

FIG. 54

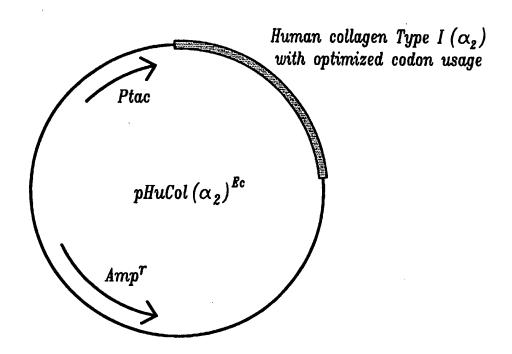


FIG. 55

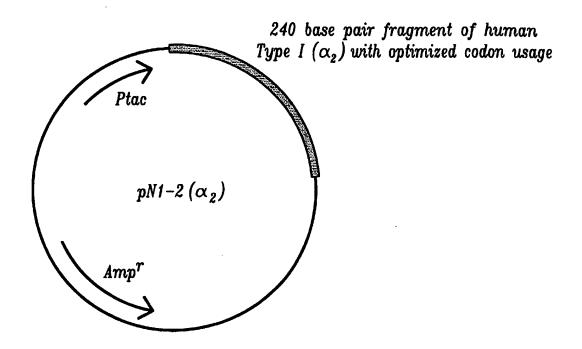


FIG. 56

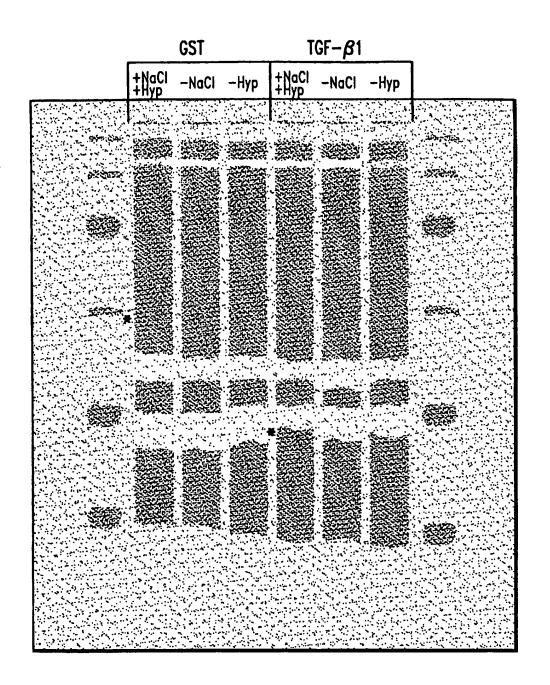


FIG. 57

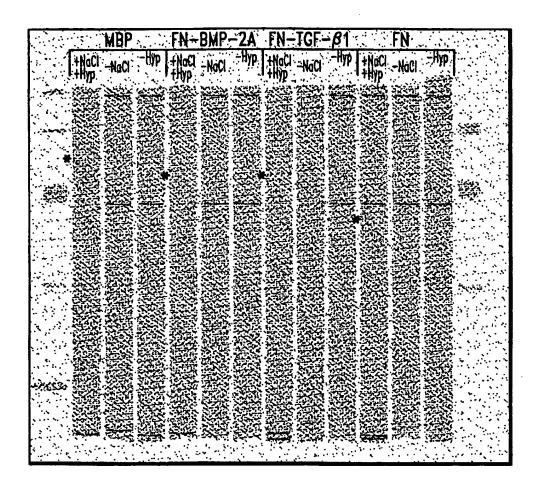


FIG. 58

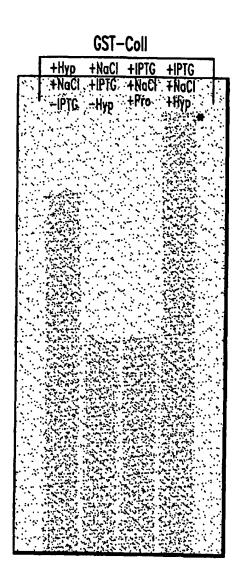


FIG. 59

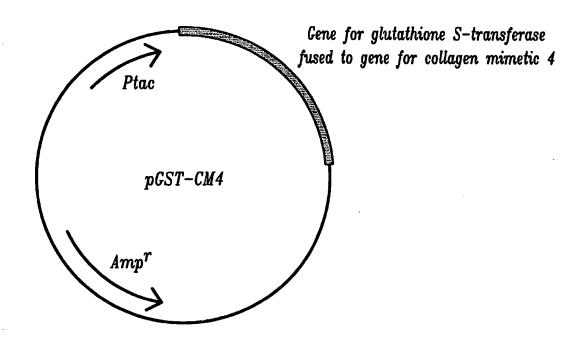
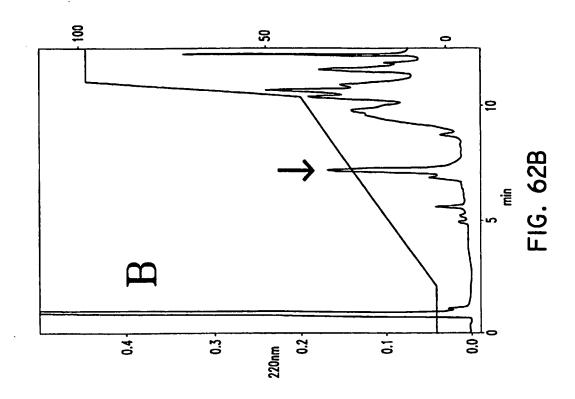
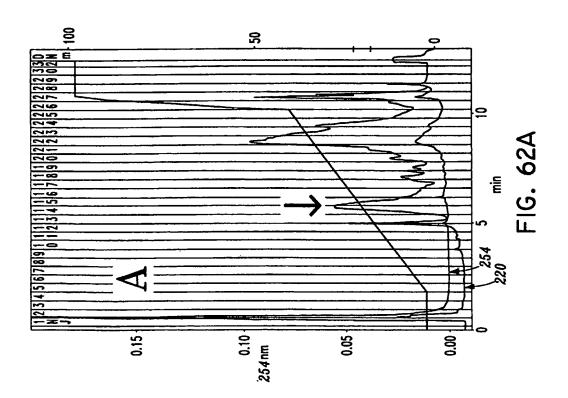


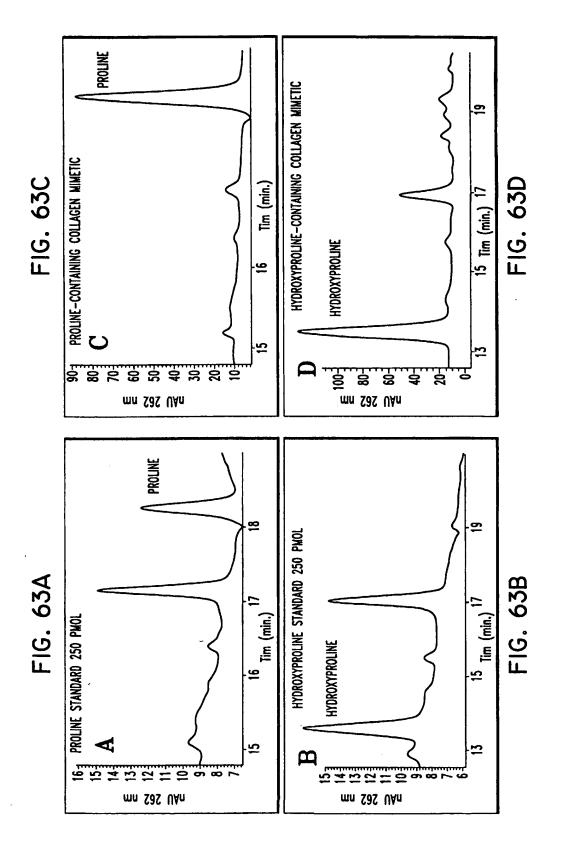
FIG. 60

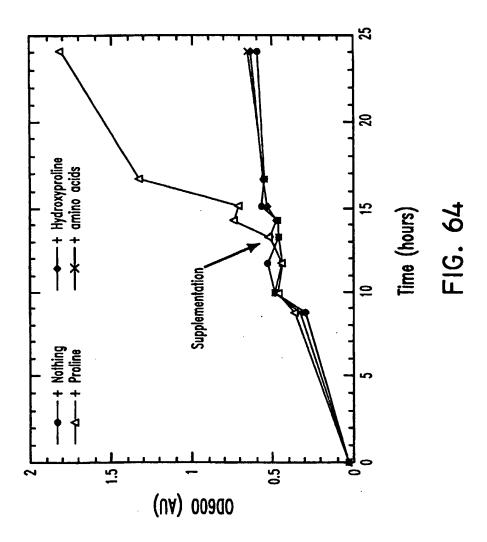
		;	9			18			27			36			45			54
5'	ATG	GGG	CTC	GCT	GGC	CCA	CCG	GGC	GAA	CCG	GGT	CCG	CCA	GGC	CCG	AAA	GGT	CCG
	М	G	L	Α	G	P	P	G	E	P	G	P	P	G	P	ĸ	G	P
		•	63			72			81			90			99			108
	CGT	GGC	GAT	AGC	GGG	CLC	GCT	GGC	CCA	CCG	GGC	GAA	CCG	GGT	CCG	CCA	GGC	CCG
		-;																
	R	G	D	S	G	L	A	G	P	P	G	E	. P	G	P	P	G	P
			117			126			135			144			153			162
	AAA	GGT										CCA						
	ĸ	G	P	R	G	α	S	G	L	A	G	P.	P	G	E	P	G	P
		;																
		•	171									198						216
	CCA	ĢGC	CCG	AAA	GGT	CCG	CGT	GGC	GAT	AGC	GGG	CTC	GCT	GGC	CCA	CCG	GGC	GAA
		-j																
	P	· G	P	K	G	P	R	G	D	S	G	L	A	G	P	P	G	E
		:	225			234			243			252			261		,,.	270
	CCG	ĠGT	CCG	CCA	GGC							GAT						
	P	·G	P	P	G	P	K	G	P	R	G	a	S	G	L	P	G	D
		:														_	_	_
		:																
	TCC	TAA	3'															
																		-
	S	*																

FIG. 61









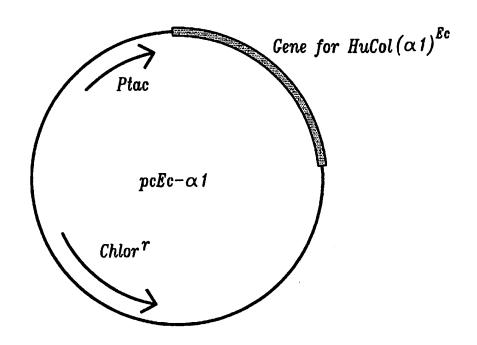


FIG. 65

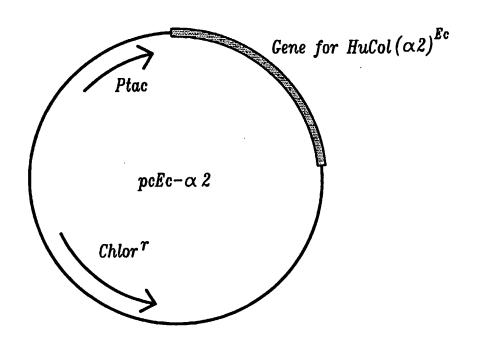


FIG. 66

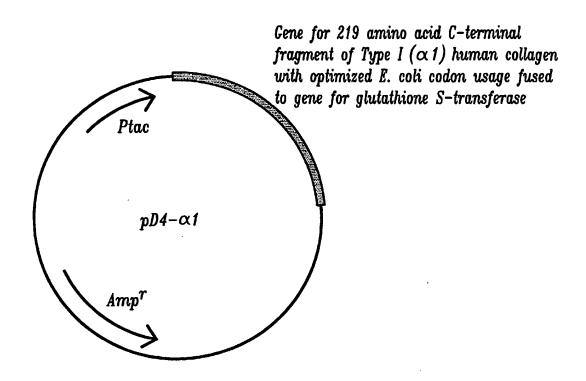


FIG. 67

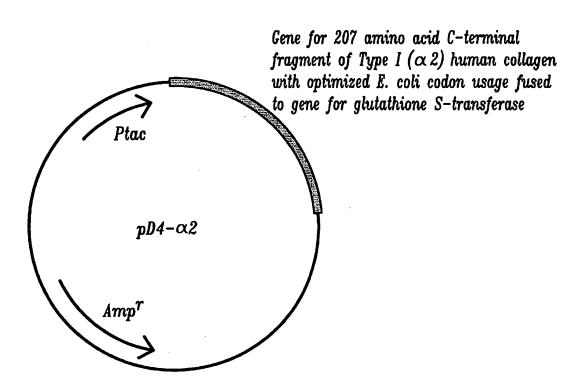
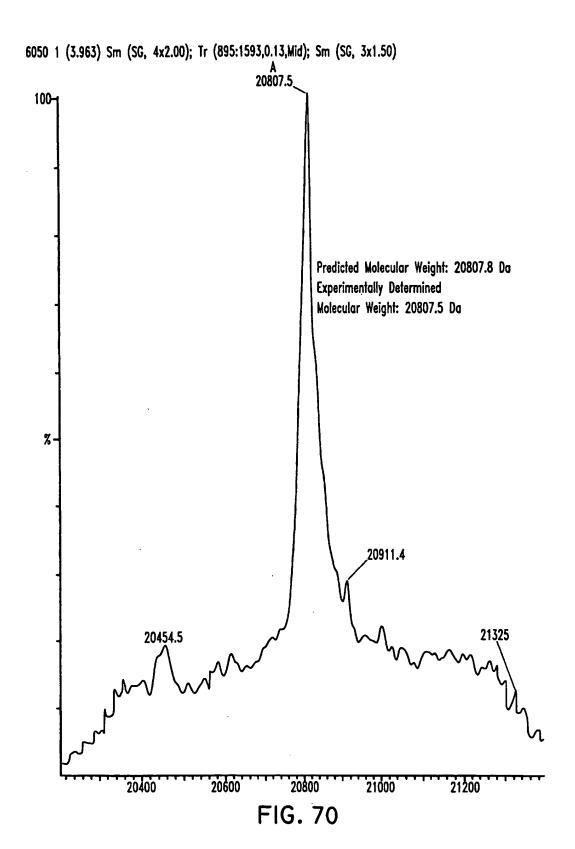


FIG. 68

Protein Sequence of the First 13 amino acids of D4-α1. Predicted From the DNA Sequence: H₂N-Gly-Pro-Pro-Gly-Leu-Ala-Gly-Pro-Pro-Gly-Glu-Ser-Gly

Experimentally-Determined Protein Sequence of the First 13 amino acids of D4-α1:

H₂N-Gly-Hyp-Hyp-Gly-Leu-Ala-Gly-Hyp-Hyp-Gly-Glu-Ser-Gly



c	~~~	9	<i>(:</i> CC	CCT	18	ccc	CCC	27 CCT	ccc	CCT	36	¥CC	CCT	45 CCT	CAA	CCC	54 GCG	ccc	ССТ
ccc	C7.7	69	≱CC	CCA	78	ccc	GAC	ggr Ggr	AGC	CCG	96 CGG	GCC.	AAA	105 GGG	TAD	CGT	114 GGT	GAA	ACC
ccc	OM																	_,,	
CCC	ccc	129 GCC	OGC	ccc	CCC	CCT	GCA	CCC	GGC	GCG	CCC	GGT	GCC	762	GGC	CCG	GTG	GGC	CCG
		120			199			207			216			225			23.1		
GCG	GGC	AAA	VCC	GGT	GAT	CGT	GGT	GAG	ACC	GGT	ccc	GCG	GGC	CCC	GCC	GCT	CCG	GTG	GGC
		249			253			267			276			285			294		
CCA	GCG	GGC	GCC	CGI	CCC	CCG	GCC	GCT	CCG	CAG	GGC	CCG	CGG	GGT	GAC	AAA	GGT	GAA	ACG
		309	222	<b></b>	313	CCC	» (IVI)	327	CCC	C) G	336		~~~	345			354 CAG		
CCG	GCC	369 CCG	CCG	GGC	378 201	CCG	CCT	387 GAA	CAG	GCT	396 CCG	TCC	GGA	405 GCC	AGC	GGG	414 CCG	GCG	ccc
•				•														0.0	
CCA	CGC	GGT	CCG	င်ငဒ	GGC	AGC	CCC	GGC	CCG	ccc	CCC	ааа	GAC	CGT	CTG	AAC	GGT	CTG	CCC
		489			493			507			5).6			525			534		
CCC	CCG	ATC	GGC	CCG	ĆCC	GGC	CCA	CGC	CCC	CGC	ACC	CCT	Gat	CCC	GGT	CCC	GTG	GGT	CCC
000	<b></b>	549		200	558			567			576			585			594		
المال	CCC																GAC	TTC	AGC
יאדני	CTG	609 CCG	CaG	cce	63.8	CAG	CYC	627	CCC	רזיר	636	ccc	CCT	645	መእሮ	(O) C	654 CCT	000	
	0																	GCG	TAA
					0/3			057			596			705	•••		714		

FIG. 71

70	20	30	40	50	60
MGPFGLAGPP	GESGREGAPG	REGSPGREGS	PGAKGDRGET	GPAGPPGAPG	APGAPGPVGP
70	30	90	100	- 110	120
AGKSGDRGET	GPAGPAGPVG	Pagargpagp	QGPRGDKGET	GEOGORGIKG	HRGFSGLQGP
130	140	150	160	170	180
PGPPGSPCEQ	GI'SGASGPAG	PRGPPGSAGA	PGKDGLNGLP	GPIGPPGPRG	RTGDAGPVGP
190	200	210	220	230	. 240
PGPPGPPCPP	GPPSAGEDES	FLPQPFQEKA	HEGGEYYRA*		

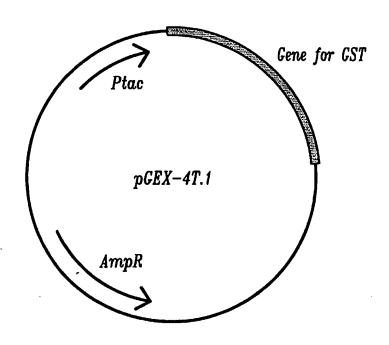


FIG. 73

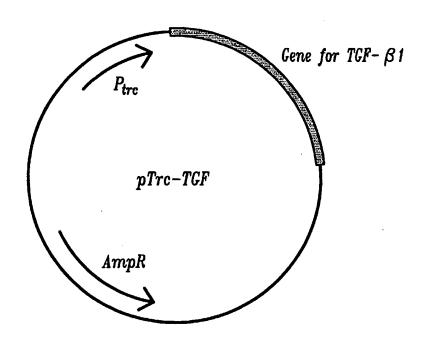


FIG. 74

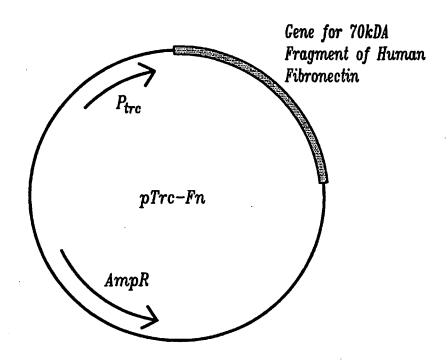


FIG. 75

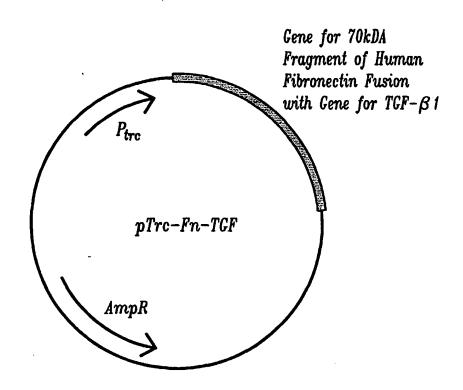


FIG. 76

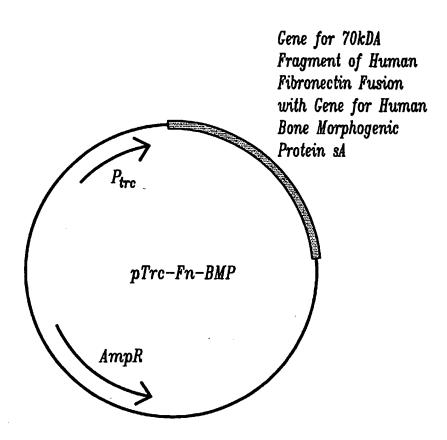


FIG. 77

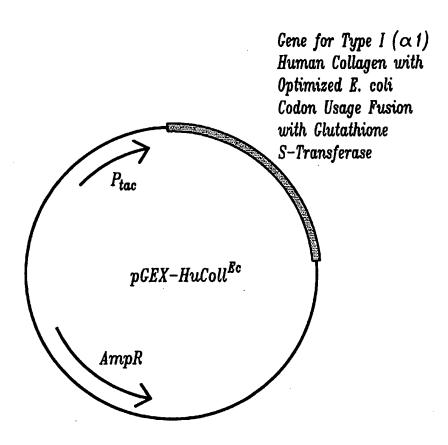


FIG. 78

FIG. 79

FIG. 80

#### Oligo N4-1

5'GGAATTCTCCCATGGGCCGGCCGGGTCTGGCGGGCCCTCCGGGTGAAAGCGGTCGTGA AGGCGCGCGGGTGCCGAAGGCAGCCCAGGCCGCGAC

#### Oligo N4-2

3'CTTCCGTCGGGTCCGGCGCTGCCATCGGGCCCCCGGTTTCCCCTAGCACCACTTTGGCC GGGCCGCCCGGGGGGCCCACGTGGCATTATTCGAACCC

#### Oligo N4-3

5'GGAATTCGGTGCACCGGGCGCGCGGGCGGGCAAA AGCGGTGATCGTGGCGAGACCGGTCCGGCGGGC

## Oligo N4-4

3'CTCTGGCCAGGCCGGGCCGGGCCAGGCCACCGGGTCGCCGGGCACCGGGCC GGCCAGGCGTCCCGGGCGCCATTATTCGAACCC

# This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

# **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

BLACK BORDERS

IMAGE CUT OFF AT TOP, BOTTOM OR SIDES

FADED TEXT OR DRAWING

BLURRED OR ILLEGIBLE TEXT OR DRAWING

SKEWED/SLANTED IMAGES

COLOR OR BLACK AND WHITE PHOTOGRAPHS

GRAY SCALE DOCUMENTS

LINES OR MARKS ON ORIGINAL DOCUMENT

REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY

IMAGES ARE BEST AVAILABLE COPY.

OTHER:

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.